Soil Moisture Active Passive (SMAP) Project

Algorithm Theoretical Basis Document (ATBD)

SMAP Level 4 Surface and Root Zone Soil Moisture (L4_SM) Data Product

Initial Release, v.1

Rolf Reichle[1], Wade Crow[2], Randal Koster[1], John Kimball[3], and Gabrielle De Lannoy[1,4,5]

1 – Global Modeling and Assimilation Office, NASA/GSFC, Greenbelt, MD, USA
2 – Hydrology and Remote Sensing Lab, USDA/ARS, Beltsville, MD, USA
3 – Flathead Lake Biological Station, University of Montana, Polson, MT, USA
4 – Universities Space Research Association, Columbia, MD, USA
5 – Ghent University, Ghent, Belgium

Corresponding author: Rolf Reichle
NASA Goddard Space Flight Center
Code 610.1
Greenbelt Rd
Greenbelt, MD, 20771
USA
Email: Rolf.Reichle@nasa.gov
Phone: +1-301-614-5693

Algorithm Theoretical Basis Documents (ATBDs) provide the physical and mathematical descriptions of the algorithms used in the generation of science data products. The ATBDs include a description of variance and uncertainty estimates and considerations of calibration and validation, exception control and diagnostics. Internal and external data flows are also described.

The SMAP ATBDs were reviewed by a NASA Headquarters review panel in January 2012 and are currently at Initial Release, version 1. The ATBDs will undergo additional updates after the SMAP Algorithm Review in September 2013.

SMAP Level 4 Surface and Root Zone Soil Moisture (L4_SM) Data Product
Algorithm Theoretical Basis Document

EXECUTIVE SUMMARY

Rolf Reichle*, Wade Crow, Randal Koster, John Kimball, and Gabrielle De Lannoy

*Email: Rolf.Reichle@nasa.gov
Phone: +1-301-614-5693

The Soil Moisture Active Passive (SMAP) mission (planned launch in October 2014) measures land surface microwave emission (or brightness temperature; at 1.41 GHz) and radar backscatter (at 1.26 GHz and 1.29 GHz), thus providing information on surface soil moisture (top 5 cm of the soil column) and on the freeze/thaw state of the land surface. The main objectives of the SMAP Level 4 Surface and Root Zone Soil Moisture (L4_SM) product are

(i) to provide estimates of *root zone* soil moisture (defined here nominally as soil moisture in the top 1 m of the soil column) based on SMAP observations, and

(ii) to provide a global surface and root zone soil moisture product that is spatially and temporally complete.

Obtaining root zone soil moisture is important for several of the key applications targeted by SMAP. The L4_SM algorithm uses an ensemble Kalman filter (EnKF) to merge SMAP observations with soil moisture estimates from the NASA Catchment land surface model. The Catchment model describes the vertical transfer of soil moisture between the surface and root zone reservoirs and will be driven with observation-based surface meteorological forcing data, including precipitation, on a global 9 km Earth-fixed grid with a 20 min model time step. The EnKF considers the respective uncertainties of each component estimate and, if provided with properly calibrated uncertainty inputs, yields a product that is superior to satellite or land model data alone. Error estimates for the L4_SM product are also generated.

The baseline L4_SM algorithm assimilates the following three SMAP products: (i) 9 km downscaled brightness temperature (downscaled from 36 km by the L2_SM_AP algorithm based on high-resolution radar backscatter measurements), (ii) 36 km brightness temperature (from L1C_TB) when and where downscaled brightness temperature is not available, and (iii) freeze-thaw observations (from L3_FT_A). Analysis updates are computed every three hours (at 0z, 3z, …, 21z) using the available SMAP products. The baseline L4_SM algorithm assimilates brightness temperature Z-scores (standardized anomalies with respect to a long-term climatology that resolves the seasonal and diurnal cycles). Freeze-thaw observations are assimilated with a rule-based approach similar to existing snow cover assimilation methods. Three optional extensions to the L4_SM algorithm are considered: (i) A dynamic bias estimation module may be used in combination with or instead of the a priori scaling approach, (ii) the filter may be replaced with a smoother, which is particularly attractive for reprocessing, and (iii) an adaptive filter module may be used to refine estimates of model and observation error standard deviations.

The L4_SM product provides a variety of geophysical fields at 3 hour time resolution on the global 9 km modeling grid, along with the assimilated lower-level SMAP observations and related instantaneous model and analysis fields. Surface and root zone soil moisture will be output in percentile units. Moreover, surface soil moisture will be output in volumetric units (m^3 m^{-3}) consistent with the climatology of the L2_SM_AP product (only for locations where L2_SM_AP is available). L4_SM surface and root zone soil moisture estimates will be validated to an RMSE requirement of 0.04 m^3 m^{-3} after removal of the long-term mean bias (identical to Level 2 soil moisture product validation and excluding regions of snow and ice, frozen ground, mountainous topography, open water, urban areas, and vegetation with water content greater than 5 kg m^{-2}). Research outputs (not validated) include the surface meteorological forcing fields, land surface fluxes, soil temperature and snow states, runoff, and error estimates (derived from the ensemble).

Based on the assimilation of surface soil moisture retrievals from the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) and the Advanced Scatterometer (ASCAT) and based on Observing System Simulation Experiments (OSSEs), it is expected that the L4_SM soil moisture estimates will meet the error requirements. It was further demonstrated that the assimilation of satellite soil moisture retrievals from AMSR-E and the correction of the precipitation forcing towards gauge- and satellite-based estimates contribute significant and largely independent amounts of information. Because SMAP operates at L-band whereas AMSR-E and ASCAT retrievals are based on X-band (10.7 GHz) and C-band (5.3 GHz), respectively, it is expected that SMAP observations further improve soil moisture assimilation estimates.

Pre-launch calibration will be conducted primarily with observations from the Soil Moisture Ocean Salinity (SMOS) mission and with OSSEs. Post-launch, the L4_SM soil moisture output will be validated against in situ observations from targeted field campaigns, SMAP core sites, and sparse networks. Moreover, the L4_SM product will be evaluated using internal diagnostics from the assimilation system (such as the observation-minus-model residuals and the soil moisture increments) in conjunction with high-quality, independent precipitation observations.

TABLE OF CONTENTS

1. Introduction

1.1 Mission Background and Science Objectives

The National Research Council's (NRC) Decadal Survey, *Earth Science and Applications from Space: National Imperatives for the Next Decade and Beyond*, was released in 2007 after a two year study commissioned by NASA, NOAA, and USGS to provide prioritization recommendations for space-based Earth observation programs [*Space Studies Board,* 2007]. Factors including scientific value, societal benefit and technical maturity of mission concepts were considered as criteria. SMAP data products have high science value and provide data towards improving many natural hazards applications. Furthermore SMAP draws on the significant design and risk-reduction heritage of the Hydrosphere State (Hydros) mission [*Entekhabi et al.*, 2004]. For these reasons, the NRC report placed SMAP in the first tier of missions in its survey. In 2008 NASA announced the formation of the SMAP project as a joint effort of NASA's Jet Propulsion Laboratory (JPL) and Goddard Space Flight Center (GSFC), with project management responsibilities at JPL. The target launch date is October 2014 [*Entekhabi et al.*, 2010b].

The SMAP science and applications objectives are to:

- Understand processes that link the terrestrial water, energy and carbon cycles;
- Estimate global water and energy fluxes at the land surface;
- Quantify net carbon flux in boreal landscapes;
- Enhance weather and climate forecast skill;
- Develop improved flood prediction and drought monitoring capability.

1.2 Measurement Approach

Table 1 is a summary of the SMAP instrument functional requirements derived from its science measurement needs. The goal is to combine the attributes of the radar and radiometer observations (in terms of their spatial resolution and sensitivity to soil moisture, surface roughness, and vegetation) to estimate soil moisture at a resolution of 10 km, and freeze-thaw state at a resolution of 1-3 km.

The SMAP instrument incorporates an L-band radar and an L-band radiometer that share a single feedhorn and parabolic mesh reflector. As shown in Figure 1 the reflector is offset from nadir and rotates about the nadir axis at 14.6 rpm (nominal), providing a conically scanning antenna beam with a surface incidence angle of approximately 40°. The provision of constant incidence angle across the swath simplifies the data processing and enables accurate repeat-pass estimation of soil moisture and freeze/thaw change. The reflector has a diameter of 6 m, providing a radiometer 3 dB antenna footprint of 40 km (root-ellipsoidal-area). The real-aperture radar footprint is 30 km, defined by the two-way antenna beamwidth. The real-aperture radar and radiometer data will be collected globally during both ascending and descending passes.

Scientific Measurement Requirements	Instrument Functional Requirements
Soil Moisture: ~±0.04 $m^3 m^{-3}$ volumetric accuracy (1-sigma) in the top 5 cm for vegetation water content ≤ 5 kg m^{-2}; Hydrometeorology at ~10 km resolution; Hydroclimatology at ~40 km resolution	L-Band Radiometer (1.41 GHz): Polarization: V, H, T_3 and T_4 Resolution: 40 km Radiometric Uncertainty*: 1.3 K L-Band Radar (1.26 and 1.29 GHz): Polarization: VV, HH, HV (or VH) Resolution: 10 km Relative accuracy*: 0.5 dB (VV and HH) Constant incidence angle** between 35° and 50°
Freeze/Thaw State: Capture freeze/thaw state transitions in integrated vegetation-soil continuum with two-day precision, at the spatial scale of land-scape variability (~3 km).	L-Band Radar (1.26 GHz and 1.29 GHz): Polarization: HH Resolution: 3 km Relative accuracy*: 0.7 dB (1 dB per channel if 2 channels are used) Constant incidence angle** between 35° and 50°
Sample diurnal cycle at consistent time of day (6am/6pm Equator crossing); Global, ~3 day (or better) revisit; Boreal, ~2 day (or better) revisit	Swath Width: ~1000 km Minimize Faraday rotation (degradation factor at L-band)
Observation over minimum of three annual cycles	Baseline three-year mission life
* Includes precision and calibration stability ** Defined without regard to local topographic variation	

Table 1. SMAP Mission Requirements

To obtain the desired high spatial resolution the radar employs range and Doppler discrimination. The radar data can be processed to yield resolution enhancement to 1-3 km spatial resolution over the 70% outer parts of the 1000 km swath. Data volume prohibits the downlink of the entire radar data acquisition. Radar measurements that allow high-resolution processing will be collected during the morning overpass over all land regions and extending one swath width over the surrounding oceans. During the evening overpass data poleward of 45° N will be collected and processed as well to support robust detection of boreal landscape freeze/thaw transitions.

The baseline orbit parameters are:

- Orbit Altitude: 685 km (2-3 days average revisit and 8-days exact repeat)
- Inclination: 98°, sun-synchronous
- Local Time of Ascending Node: 6 pm

The SMAP radiometer measures the four Stokes parameters, V, H and T_3, and T_4 at 1.41 GHz. The T_3-channel measurement can be used to correct for possible Faraday rotation caused by the ionosphere, although such Faraday rotation is minimized by the selection of the 6am/6pm sun-synchronous SMAP orbit.

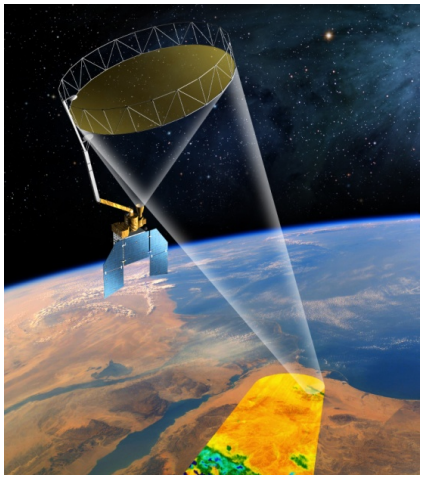

Figure 1. The SMAP observatory is a dedicated spacecraft with a rotating 6 m light-weight deployable mesh reflector. The radar and radiometer share a common feed.

At L-band anthropogenic Radio Frequency Interference (RFI), principally from ground-based surveillance radars, can contaminate both radar and radiometer measurements. Early measurements and results from the SMOS mission indicate that in some regions RFI is present and detectable. The SMAP radar and radiometer electronics and algorithms have been designed to mitigate the effects of RFI. The SMAP radar utilizes selective filters and an adjustable carrier frequency in order to tune to pre-determined RFI-free portions of the spectrum while on orbit. The SMAP radiometer will implement a combination of time and frequency diversity, kurtosis detection, and use of T_4 thresholds to detect and, where possible, mitigate RFI.

The planned SMAP data products are listed in Table 2. Level 1B and 1C data products are calibrated and geolocated instrument measurements of surface radar backscatter cross-section and brightness temperatures derived from antenna temperatures. Level 2 products are geophysical retrievals of soil moisture on a Earth-fixed grid based on Level 1 products and ancillary information; the Level 2 products are output on a half-orbit basis. Level 3 products are daily composites of Level 2 surface soil moisture and freeze/thaw state data. Level 4 products are model-derived value-added data products that support key SMAP applications and more directly address the driving science questions.

1.3 Motivation for the L4_SM Data Product

The primary SMAP measurements, land surface microwave emission at 1.41 GHz and radar backscatter at 1.26 GHz and 1.29 GHz, are directly related to *surface* soil moisture (in the top 5 cm of the soil column). Several of the key applications targeted by SMAP, however, require knowledge of **root zone soil moisture** (defined here as soil moisture in the top 1 m of the soil column), which is not directly linked to SMAP observations. The foremost objective of the SMAP Level 4 Surface and Root Zone Soil Moisture (L4_SM) product is to fill this gap and provide estimates of root zone soil moisture that are informed by and consistent with SMAP observations. Such estimates are obtained by merging SMAP observations with estimates from a land surface model in a soil moisture data assimilation system.

Product	Description	Gridding (Resolution)	Latency	
L1A_TB	Radiometer Data in Time-Order	-	12 hrs	Instrument Data
L1A_S0	Radar Data in Time-Order	-	12 hrs	
L1B_TB	Radiometer T_B in Time-Order	(36x47 km)	12 hrs	
L1B_S0_LoRes	Low Resolution Radar σ_o in Time-Order	(5x30 km)	12 hrs	
L1C_S0_HiRes	High Resolution Radar σ_o in Half-Orbits	1 km (1-3 km)	12 hrs	
L1C_TB	Radiometer T_B in Half-Orbits	36 km	12 hrs	
L2_SM_A	Soil Moisture (Radar)	3 km	24 hrs	Science Data (Half-Orbit)
L2_SM_P	Soil Moisture (Radiometer)	36 km	24 hrs	
L2_SM_AP	Soil Moisture (Radar + Radiometer)	9 km	24 hrs	
L3_FT_A	Freeze/Thaw State (Radar)	3 km	50 hrs	Science Data (Daily Composite)
L3_SM_A	Soil Moisture (Radar)	3 km	50 hrs	
L3_SM_P	Soil Moisture (Radiometer)	36 km	50 hrs	
L3_SM_AP	Soil Moisture (Radar + Radiometer)	9 km	50 hrs	
L4_SM	Soil Moisture (Surface and Root Zone)	9 km	7 days	Science Value-Added
L4_C	Carbon Net Ecosystem Exchange (NEE)	9 km	14 days	

Table 2. Overview of SMAP science data products. Mean latency under normal operating conditions. The SMAP project will make a best effort to reduce these latencies.

The land surface model component of the assimilation system is driven with observation-based surface meteorological forcing data, including precipitation, which is the most important driver for soil moisture. The model also encapsulates knowledge of key land surface processes, including the vertical transfer of soil moisture between the surface and root zone reservoirs. Finally, the model interpolates and extrapolates SMAP observations in time and in space. The SMAP L4_SM product thus provides a comprehensive and consistent picture of land surface hydrological conditions based on SMAP observations and complementary information from a variety of sources. The assimilation algorithm considers the respective uncertainties of each component and yields a product that is superior to satellite or land model data alone. Error estimates for the L4_SM product are generated as a by-product of the data assimilation system.

Without root zone soil moisture estimates from the L4_SM product, SMAP would be of limited use for several key applications targeted by the mission. Fortunately, there has been substantial progress in land data assimilation over the past decade, and soil moisture data assimilation has already been demonstrated successfully with SMAP precursor datasets (section 2.2). Much of the development occurred at GSFC, which is a partner in the SMAP project. The Global Modeling and Assimilation Office (GMAO) at GSFC currently hosts an ensemble Kalman filter (EnKF) data assimilation system that has been used successfully to assimilate satellite retrievals of surface soil moisture into the NASA Catchment land surface model. When the satellite data, the model data, and the assimilation product are each compared to independent in-situ observations, the assimilated product proves superior (section 2.2), thereby validating the

assimilation system [*Reichle et al.*, 2007; *Liu et al.*, 2011; *Draper et al.,* 2012]. Most importantly, the assimilation system improves over the model-only root zone soil moisture estimates. This ATBD provides a detailed description of the SMAP L4_SM product, its algorithm, and how the product will be validated.

2. Overview and Background

2.1 Product/Algorithm Objectives

The main objectives of the SMAP L4_SM product are

(i) to provide estimates of *root zone* soil moisture (defined here nominally as soil moisture in the top 1 m of the soil column) based on SMAP observations, and
(ii) to provide a global surface and root zone soil moisture product that is spatially and temporally complete.

These objectives address two limitations of the SMAP Level 2 soil moisture products, which provide soil moisture estimates only for the surface layer (~top 5 cm of the soil) and only at times and locations where soil moisture can be observed by SMAP sensors (subject to orbit and land surface characteristics).

2.2 Historical Perspective

There has been considerable progress in the methodological development of soil moisture data assimilation [*Walker and Houser*, 2001; *Reichle et al.*, 2002a,b; *Margulis et al.*, 2002; *Reichle and Koster*, 2003; *Crow and Wood*, 2003; *Seuffert et al.*, 2003; *Crow and van Loon*, 2006; *De Lannoy et al.*, 2007; *Dunne and Entekhabi*, 2006; *Pan and Wood*, 2006; *Zhou et al.*, 2006], with ensemble-based Kalman filtering and smoothing algorithms emerging as the methods of choice. These developments were initially based on assimilation experiments with synthetic soil moisture retrievals and field-scale studies because global satellite observations of soil moisture had been lacking. Recently, a number of such data sets have become available, including soil moisture products from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E; since 2002) [*Njoku*, 2011, *Owe et al.*, 2008], the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI; since 1997] [*Gao et al.*, 2006, *Owe et al.,* 2008], Windsat [*Li et al.*, 2007], and the historic Scanning Multichannel Microwave Radiometer (SMMR;1978-1987) [*Owe et al.*, 2008]. These products are based on C- and X-band passive microwave observations with an effective sensing depth of roughly 1 cm. Soil moisture retrievals can also be obtained from (active) radar microwave measurements, including those from the European Remote Sensing satellites (ERS-1/2) [*Wagner et al.,* 2007], and the Advanced Scatterometer (ASCAT) [*Bartalis et al.*, 2008]. Improved retrievals are expected from passive L-band sensors that measure moisture in the top 5 cm of the soil, including SMOS [*Kerr et al.*, 2001; 2010] launched in November 2009, the NASA Aquarius mission [*Lagerloef et al.*, 2008] launched in June 2011, and, of course, the NASA SMAP mission.

Significant climatological differences have been identified between independent soil moisture estimates from in situ measurements, satellite retrievals, and model integrations of antecedent meteorological forcing data. On a global scale, neither the satellite nor the model soil moisture are more consistent with the available in situ observations, implying that presently there is no agreed climatology of global soil moisture [*Reichle et al.*, 2004]. This gap may be filled in the future through a sufficiently long record of satellite-based soil moisture observations at L-band

(such as provided by SMOS and SMAP) and validation against similarly long records of distributed in situ observations. In the meantime, the problem can be circumvented for data assimilation through a scaling approach [*Reichle and Koster*, 2004; *Drusch et al.,* 2005]. The central idea is to rescale the satellite data prior to assimilation by matching the satellite data's cumulative distribution function (cdf) to the model's climatology. When using such rescaling, the key information is encoded in the anomaly time series, which in any case is of primary interest for many important applications such as numerical weather and seasonal climate forecast initialization. Hence, long-term mean bias should be removed prior to validation (Section 4.2.4). Obviously, such validation relies on the availability of relatively long time series of satellite and in situ observations, though approaches using ergodic sampling have been developed [*Reichle and Koster*, 2004] that relax the time-series length requirement for satellite data (section 4.1.2e), allowing even short missions like SMAP to contribute effectively to Level 4 product generation.

It has been demonstrated that the assimilation of satellite retrievals of surface soil moisture into a land model does in fact yield superior estimates of soil moisture conditions when compared to model or satellite estimates alone. *Reichle and Koster* [2005], *Drusch* [2007], and *Reichle et al.* [2007] demonstrated this property for large-scale soil moisture fields based on the assimilation of retrievals from SMMR, TMI, and AMSR-E, respectively. Recently, *Liu et al.* [2011] assessed the contributions of precipitation and soil moisture observations to soil moisture skill in a land data assimilation system. Relative to baseline estimates from the Modern Era Retrospective-analysis for Research and Applications (MERRA), the study investigated soil moisture skill derived from (i) model forcing corrections based on large-scale, gauge- and satellite-based precipitation observations and (ii) assimilation of surface soil moisture retrievals from AMSR-E. Soil moisture skill (defined as the anomaly time series correlation coefficient R) was assessed using in situ observations in the continental United States at 37 single-profile sites within the Soil Climate Analysis Network (SCAN) for which skillful AMSR-E retrievals are available and at four USDA Agricultural Research Service ("CalVal") watersheds with high-quality distributed sensor networks that measure soil moisture at the scale of land model and satellite estimates (section 4.2.4). The average skill of AMSR-E retrievals is R=0.42 versus SCAN (Figure 2) and R=0.55 versus CalVal measurements (not shown). The skill of MERRA surface and root zone soil moisture is R=0.43 and R=0.47, respectively, versus SCAN measurements (Figure 2). MERRA surface moisture skill is R=0.56 versus CalVal measurements (not shown). Most importantly, Figure 2 shows that adding information from precipitation observations increases (surface and root zone) soil moisture skills by $\Delta R \sim 0.06$, that assimilating AMSR-E retrievals increases soil moisture skills by $\Delta R \sim 0.08$, and that adding information from both sources increases soil moisture skills by $\Delta R \sim 0.13$. The result demonstrates that precipitation corrections and assimilation of satellite soil moisture retrievals contribute important and largely independent amounts of information.

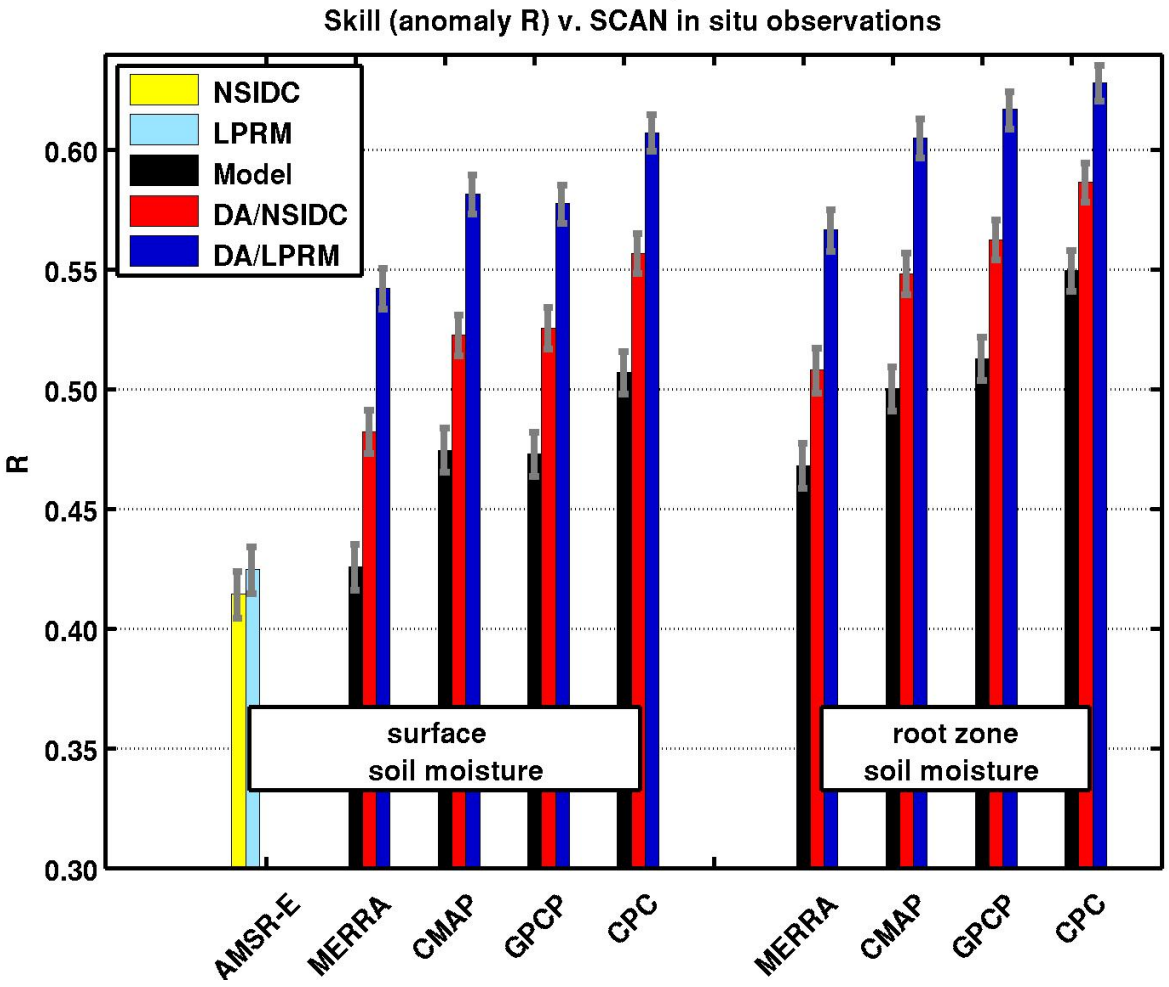

Figure 2. Average time series correlation coefficient R with SCAN in situ surface and root zone soil moisture anomalies for estimates from two AMSR-E retrieval datasets (NSIDC and LPRM), the Catchment model forced with four different precipitation datasets (MERRA, CMAP, GPCP, and CPC), and the corresponding data assimilation integrations (DA/NSIDC and DA/LPRM). Average is based on 37 SCAN sites for surface and 35 SCAN sites for root zone soil moisture (see Figure 18 in section 4.2.4). Error bars indicate approximate 95% confidence intervals.

Note that the improvements through data assimilation that are shown in Figure 2 are affected by errors in the validating in situ measurements and the fact that the in situ measurements are point-scale observations. By contrast, satellite, model, and assimilation estimates refer to horizontally distributed soil moisture footprints of about 50 km linear scale. Given the in situ errors and the scale mismatch, the maximum possible R value, even with perfect estimates of large scale soil moisture, is less than 1, and thus the actual improvement associated with data assimilation is likely to be larger than that suggested by the figure.

In a study by *Draper et al.* [2012], ASCAT and AMSR-E surface soil moisture retrievals were assimilated separately or together in the same assimilation system, and the results were validated against SCAN observations and Murrumbidgee in situ measurements. Figure 3 shows significant skill gains for surface and root zone soil moisture estimates through assimilation of either active (ASCAT) or passive (AMSR-E) data. The successful use of the active microwave

retrievals by *Draper et al.* [2012] is especially reassuring because it suggests that active measurements contain information about soil moisture. This information is captured in the L4_SM product through the use of downscaled brightness temperatures from the L2_SM_AP product.

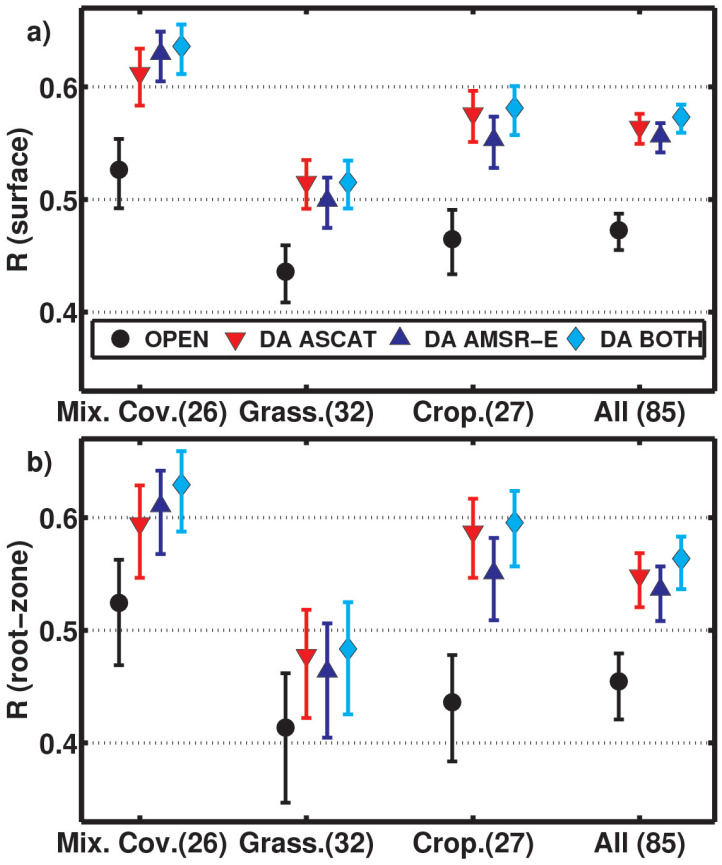

Figure 3. Mean skill for (a) surface and (b) root-zone soil moisture from the open loop model and the data assimilation (DA) of ASCAT, AMSR-E, and both, averaged across each land cover class, with 95% confidence intervals. The number of sites in each land cover class is given in the axis labels. Skill is based on all non-frozen days in the experiment period (Jan 2007 – May 2010).

The skill improvements documented in Figures 2 and 3 suggest that the assimilation system is adequately calibrated, although not necessarily optimal. Additional gains in skill may be possible through further tuning of the model and observation error parameters that are inputs to the assimilation system, for example through adaptive filtering (section 4.1.2f). Furthermore, the potential gains from assimilation of SMAP observations (as opposed to AMSR-E or ASCAT retrievals) should be larger because the input SMAP data are expected to be superior in terms of sampled depth and coverage (see section 4.1.4 and Table 5). Note, however, that the skill improvement associated with SMAP could also be smaller if future model estimates become even more skilful, for example through better precipitation estimates from improved atmospheric modeling and data assimilation systems.

There is a relatively small number of research groups with significant experience in large-scale soil moisture data assimilation, including (in alphabetical order) US-based teams at

Colorado State University [*Jones et al.*, 2004], the Massachusetts Institute of Technology [*Entekhabi et al.*, 2004; *McLaughlin et al.*, 2006], NASA/GSFC [*Reichle et al.*, 2009a; *Kumar et al.*, 2008; *Rodell et al.*, 2003], NOAA/NCEP [Mitchell *et al.*, 2004], NOAA/NESDIS [*Zhan et al.*, 2006], Princeton University [*Wood et al.*, 2008], the University of California, Los Angeles [*Margulis et al.*, 2003], and the USDA Hydrology and Remote Sensing Lab [*Crow and Zhan,* 2007]. Foreign (non-US-based) research groups with substantial soil moisture assimilation experience include teams at weather centers (Deutscher Wetterdienst, Environment Canada, the European Centre for Medium-Range Weather Forecasting, Meteo France), the Centre d'Etudes Spatiales de la Biosphère (Toulouse, France), and Monash University, Melbourne, Australia.

Among US-based groups, the GMAO and the Hydrological Sciences Branch at NASA/GSFC have probably the highest concentration of experts and the longest experience in soil moisture data assimilation. Their Goddard Earth Observing System Model, Version 5 (GEOS-5) and Land Information System (LIS) software suites [*Reichle et al.*, 2009a; *Kumar et al.*, 2008a] are among the most mature land data assimilation systems available and are suitable candidates for the SMAP L4_SM algorithm. The basic ensemble-based GEOS-5 data assimilation modules have recently been integrated into LIS [*Kumar et al.*, 2008a]. NOAA (NCEP and NESDIS) and the Air Force Weather Agency are in the process of adopting LIS as their operational software framework for land data assimilation.

2.3 Instrument/Product Characteristics

2.3.1 Instrument/Calibration Aspects (affecting product)

The baseline L4_SM product is derived from the downscaled (9 km) brightness temperatures provided with the Level 2 Radar and Radiometer Soil Moisture (L2_SM_AP) product, the 36 km brightness temperatures observations in the Level 1C Radiometer Brightness Temperature (L1C_TB) product, and the freeze-thaw observations in the Level 3 Freeze/Thaw State (L3_FT_A) product. See corresponding ATBDs for instrument and calibration aspects of these inputs (section 8).

2.3.2 Data Product Characteristics

This section provides a summary of the SMAP L4_SM product specifications. See section 6 (Tables 6 and 7) for more detailed prototype data product specifications.

2.3.2a Geophysical parameters

The SMAP L4_SM data product includes the following components:

(i) **Surface** soil moisture (0-5 cm vertical average):
   a. In units of volumetric soil moisture ($m^3 m^{-3}$) consistent with the climatology of the L2_SM_AP product (only for locations where the L2_SM_AP product is available; SFMC_L2CLIM in Table 6a).

b. In percentile units[1] (SFMC_PRCNTL in Table 6a).

(ii) Subsurface (or **"root zone"**) soil moisture (0-100 cm vertical average) in percentile units1 (RZMC_PRCNTL in Table 6a).

(iii) Additional research products (not validated), including surface meteorological forcing variables, surface soil temperature, evaporative fraction, net radiation, etc. and **error estimates** for select output fields that are produced internally by the L4_SM algorithm (section 2.3.2e).

The soil moisture output in percentile units is expected to have the greatest scientific and operational value, and its use will be encouraged. A user familiar with the variations of soil moisture in a given applications context can directly transform the percentile product into application-specific data [*Entekhabi et al.,* 2010a]. Information on the soil moisture climatology of the Catchment land surface model will be provided as part of the ancillary data (Table 7a). This information will permit users to derive surface and root zone soil moisture in volumetric units that are consistent with the Catchment model climatology[2], thereby enabling water balance calculations in connection with the additional research products (item (iii) above). As mentioned in section 2.2, the climatology of the land surface model differs from that of the retrievals for a number of reasons. Providing soil moisture in units of percentiles is most consistent with the authors' belief that the global climatology of root zone soil moisture is unknown at present, and that imposing the Catchment model climatology would be an arbitrary choice. Should the community establish a consensus climatology prior to the final SMAP data product reprocessing, the soil moisture data product will be provided in (absolute) volumetric units.

2.3.2b Spatial resolution, posting, and coverage

All L4_SM geophysical parameters will be derived at a resolution of 9 km and posted on the SMAP Earth-fixed global grid with 9 km spacing (Table 6). The 9 km grid will be consistent with that of the EASE grid of the other SMAP products. It is anticipated that spatial sub-setting tools for the L4_SM outputs will be made available by the Distributed Active Archive Center (DAAC) charged with the archival and distribution of SMAP data. In July 2011 the National Snow and Ice Data Center (http://nsidc.org/) was selected as the DAAC for the SMAP data products (except SMAP radar data products).

2.3.2c Temporal resolution and sampling

Three basic time steps are involved in the generation of the L4_SM product:

1) the land model computational time step (20 min, section 4.1.2),

[1] For example, a surface soil moisture value for a given time and location lying in the 95th percentile of surface soil moistures produced by the land model at that location, across time for the given climate, will be given a value of 95%.

[2] Soil moisture output in volumetric units and consistent with the Catchment model climatology could be provided as part of the L4_SM data stream, but this would greatly increase the risk of inappropriate use of the data.

2) the EnKF analysis update time step, and
3) the reporting (or output) time step for the instantaneous and time average geophysical fields that are stored in the L4_SM data product.

The available SMAP observations will be assimilated in an EnKF analysis update step at the nearest 3-hourly analysis time (0z, 3z, …, and 21z). A broad variety of geophysical parameters will be provided as 3-hourly averages between these update times (Table 6a). Moreover, instantaneous (forecast and analysis) soil moisture and temperature estimates will be provided along with the assimilated observations (Table 6b). These snapshots are nominally for 0z, 3z, …, or 21z. It is anticipated that time averaging tools (to daily and monthly estimates) will be provided by the DAAC.

Upscope options: The data assimilation system produces global soil moisture estimates at every land model time step (20 min). Provided sufficient computational resources are available, hourly EnKF update steps could be performed. Provided sufficient storage capacity is available in the processing system and the DAAC, the 3-hour average output of geophysical parameters could be replaced with hourly averages.

2.3.2d Latency

After the 3-month in-orbit checkout (IOC) period (see *SMAP Level 1 Mission Requirements and Success Criteria*, section 8), the L4_SM product will be produced within 7 days of satellite data acquisition (mean latency under normal operating conditions). This target latency is motivated by today's typical operational schedule for seasonal climate forecast production. Seasonal climate forecasts are initialized with land surface conditions that rely on gauge-based pentad (5-day) average precipitation observations and thus already have a latency of about 5 days. Based on the availability of the input pentad precipitation observations (currently once every five days), we plan to deliver output once every five days covering data valid for the interval from present-minus-seven-days to present-minus-two-days. This schedule may be adjusted according to the release schedule of the input precipitation observations that will be available after launch (section 4.1.3). Note that the latency of the L4_SM product is at least that of the lower-level SMAP input products plus processing time.

Delivery of the *validated* L4_SM product will begin after the Calibration/Validation phase as specified in the *SMAP Level 1 Mission Requirements and Success Criteria* (section 8). The Calibration/Validation phase for Level 4 products covers the first twelve months after IOC.

2.3.2e Error estimates

The data assimilation system weighs the relative errors of the assimilated lower-level product (i.e., radiance or retrieval) and the land model forecast. Estimates of the error of the assimilation product are dynamically determined as a by-product of this calculation. How useful these error estimates are depends on the accuracy of the input error parameters and needs to be determined through validation (section 4.2.4). The target accuracy of the assimilated lower-level products is discussed in the corresponding ATBDs (section 8). The error estimates of the land surface

model are discussed in sections 4.13 and 4.1.4. The required input error parameters are discussed in sections 4.1.2 and 4.1.3e.

Again, each instantaneous land model field automatically comes with a corresponding instantaneous error field which will be provided for select variables. The relevant outputs are listed in Table 6b. Specifically, the error estimates are derived from the ensemble standard deviation of the analyzed fields. For soil moisture, the ensemble standard deviation is first computed from the analysis ensemble in volumetric units ($m^3\ m^{-3}$) (that is, consistent with the Catchment model climatology). Thereafter, the soil moisture values associated with plus and minus one standard deviation from the analysis (ensemble mean) soil moisture are output in percentile units. (Note that because of the non-linearity of the transformation between volumetric units and percentiles, this procedure is different from providing the ensemble standard deviation of soil moisture in percentile units.) For temperatures, the ensemble standard deviation will be provided directly in units of Kelvin. These error estimates will vary in space and time.

3. Physics of the Problem

3.1 System Model

At the heart of a land surface data assimilation system is a land surface model that monitors the evolution of soil moisture, snow, and temperature states as they respond to meteorological drivers such as rainfall and incident radiation. A land model's key strength is its reliance on conservation principles known to operate in nature. In essence, a land model is designed to conserve both water (converting precipitation inputs into evaporation, runoff, and storage change) and energy (converting incident radiation into outgoing radiation, latent heat flux, sensible heat flux, storage change, and other miscellaneous terms). Given realistic forcing, these conservation principles ensure at least some first-order reliability in the simulation products – when it rains, for example, the modeled soil will typically get wetter.

In the context of SMAP, the land model indeed provides an invaluable additional benefit: it provides a means for producing soil moisture estimates at levels below the 0-5 cm surface layer that is directly sampled by the SMAP instrument. The land model includes physically-based parameterizations for transporting moisture between the near-surface soil and, say, the root zone, which can nominally be considered to extend to a meter below the surface. A single "root zone" depth is chosen here to make the SMAP product more straightforward; in nature, the depths tapped by roots vary with vegetation type. In the course of the data assimilation process, the subsurface transport formulations in the land model (along with the subsurface assimilation updates, section 4.1.2) effectively advects SMAP-based surface soil moisture information into deeper soil levels. This deeper soil moisture product component is of critical importance to many scientific applications (agriculture, climate forecasting, etc.). In fact, for most applications, deeper soil moisture is arguably more important than the 0-5 cm surface soil moisture.

Along with this benefit comes a cost. Using a land model in a global data assimilation context as part of SMAP requires a substantial amount of ancillary data. In particular, the land model requires global fields of meteorological forcing data at suitable temporal and spatial resolutions, surface properties and parameters (hydraulic conductivity characteristics, vegetation conductance characteristics, etc.), some of which may vary seasonally, and initial model states. Further discussion of these ancillary data needs is provided in section 4.1.3.

3.2 Radiative Transfer and Backscatter

The L4_SM algorithm relates soil moisture from a land model to the microwave brightness temperature and backscatter observed by SMAP. The physics of microwave radiative transfer and backscatter are described in the SMAP Level 1-3 ATBDs (section 8).

3.3 Parameter and Model Uncertainties

Modeling uncertainties include errors in the meteorological datasets used to force the land model. Moreover, any land model used will have a wealth of limitations in its physical formulations, initialization, and assigned parameter values. As a result, simulated states invariably have errors. Consider the example of a rain event. While the soil does get wetter, the land model still has to decide, based on parameterized physics, precisely how much of the water infiltrates the soil as opposed to wetting the leaves of the vegetation or running off into rivers. Inaccuracies in the partitioning naturally lead to inaccuracies in the simulated soil moisture state. Such potential deficiencies, which must always be kept in mind when considering land model products, are ameliorated to an extent by the data assimilation process, in which the satellite data are used to steer the land model products toward independently observed values. In an ensemble-based soil moisture data assimilation system, uncertainties in the land surface model are treated explicitly in the algorithm. The success of the assimilation system thus hinges on proper specification of the input (model and observation) error parameters (see discussion in section 4.1).

A general discussion of the uncertainties in the radiative transfer and backscatter models and parameters can be found in the SMAP Level 1-3 ATBDs (section 8). The following section provides a detailed discussion of the model and parameter uncertainties of the L4_SM algorithm.

4. L4_SM Algorithm

4.1 Theoretical Description

The present version of the SMAP L4_SM ATBD includes feedback from a formal external review of the draft L4_SM algorithm and data product that was conducted in 2009 [*Reichle et al.*, 2009b].

4.1.1 Baseline and Option Algorithms Overview

The L4_SM algorithm consists of two key elements:

(i) GEOS-5 Catchment land surface and microwave radiative transfer model
(ii) GEOS-5 ensemble-based land data assimilation algorithm

The GEOS-5 Catchment land surface model is a numerical description of the water and energy transport processes at the land-atmosphere interface, augmented with a model that describes the land surface microwave radiative transfer (section 4.1.2). The GEOS-5 ensemble-based land data assimilation system (section 4.1.2) is the tool that will be used to merge SMAP observations with estimates from the land model as it is driven with observation-based surface meteorological forcing data.

The **baseline** L4_SM algorithm, described in detail in sections 4.1.2a-e, includes a soil moisture analysis based on the ensemble Kalman filter and a rule-based freeze/thaw analysis. Downscaled (9 km) brightness temperatures (L2_SM_AP) will be assimilated when and where available, supplemented with 36 km brightness temperature observations (L1C_TB; ascending and descending passes) where downscaled data are unavailable. Moreover, freeze-thaw observations (L3_FT_A) will also be assimilated. Three **optional extensions** of the baseline algorithm will be considered: smoothing, dynamic bias estimation, and adaptive filtering (section 4.1.2f).

After initialization of the system with estimates derived from a model spin-up procedure (section 4.1.3), the baseline L4_SM algorithm steps recursively through time, alternating between model forecast (FCST) and analysis (ANA) steps. Figure 4 provides an overview of one forecast and analysis cycle. The algorithm begins with a Catchment model ensemble forecast, initialized with the analysis at time t-1 and valid at time t (labeled FCST(t) in Figure 4). For each 9 km model grid cell, the forecast freeze-thaw (F/T) state is first compared to the corresponding SMAP freeze-thaw observations (aggregated to the resolution of the model forecast). If the Catchment model forecast and the SMAP observations disagree, the model states in the 9 km grid cell in question are corrected towards the observations in a freeze-thaw analysis (section 4.1.2). If the forecast and observed freeze-thaw states agree *and* indicate non-frozen conditions, the grid cell in question is included in a distributed soil moisture analysis (section 4.1.2). If the model indicates non-frozen conditions and freeze-thaw observations are not available, the grid cell is also included in the soil moisture analysis. Otherwise, the analysis step is skipped for the grid cell in question. After the analysis has been completed for all grid cells, the algorithm continues with a model forecast to time t+1, and so on.

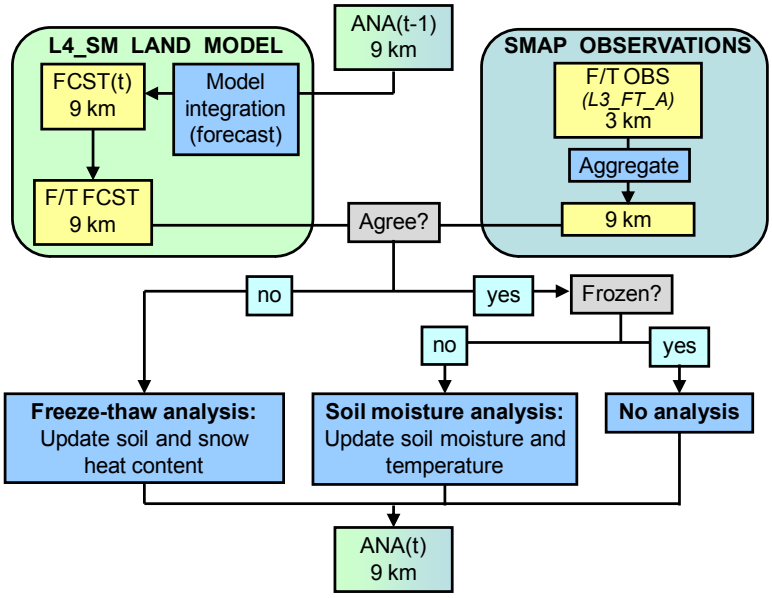

Figure 4. L4_SM algorithm overview. See Figure 7 for a flowchart of the soil moisture analysis.

Assimilation of brightness temperatures vs. radar backscatter vs. soil moisture retrievals

The assimilation of brightness temperatures requires a forward passive microwave radiative transfer model that transforms soil moisture and temperature into brightness temperature (or radiance; section 4.1.2). Such radiance assimilation is straightforward and has been described in the literature [*Reichle et al.*, 2001; *Crow and Wood*, 2002; *Seuffert et al.*, 2003]. The L2_SM_AP algorithm uses high-resolution radar backscatter measurements and 36 km brightness temperature observations to provide downscaled (9 km) brightness temperature estimates that can easily be ingested into the L4_SM algorithm. Alternatively, the L4_SM algorithm could directly assimilate radar backscatter coefficients [*Zhan et al.*, 2006; *Hoeben and Troch*, 2000], but this approach is considerably less mature and therefore not considered for the L4_SM algorithm.

Another alternative would be the assimilation of *soil moisture retrievals* from the Level 2 products (L2_SM_P, L2_SM_A, and/or L2_SM_AP). However, assimilating brightness temperatures comes with several advantages: the L4_SM product will not depend entirely on any Level 2 product and could fall back on the assimilation of L1C_TB brightness temperatures, which allows for more robust processing. Moreover, the land surface temperatures that are used in the L4_SM algorithm (for model predictions of brightness temperatures) are taken from within the data assimilation system. The baseline algorithm thus avoids a potential inconsistency that would be present if Level 2 retrievals were assimilated, because the Level 2 retrievals utilize independent estimates of surface temperature that are not necessarily consistent with those of the Level 4 system. Note that this inconsistency would still be present even if the Level 2 algorithms use ancillary soil temperature data from the NASA GMAO atmospheric analysis system because of the differences between the GMAO atmospheric analysis system and the off-line, SMAP-customized L4_SM system.

4.1.2 Mathematical Description of the Algorithm

In this section, we provide the mathematical descriptions of the key elements of the baseline and option algorithms.

4.1.2a The NASA Catchment land surface model

Model soil moisture is obtained from integrations of the NASA Catchment Land Surface Model [hereinafter Catchment model; *Koster et al*., 2000; *Ducharne et al*., 2000], which has been developed by the NASA GMAO. Although in standard practice the basic computational unit of the Catchment model is the hydrological catchment (or watershed), for SMAP we will use the Earth-fixed 9 km EASE grid (same as that of the L2_SM_AP product) to define the surface elements. The conceptual physics underlying the model, which focus on topographical variations smaller than the 9 km scale, are still important and valid for such a surface element definition.

Table 3 lists the Catchment model prognostic variables, and Figure 5 provides a simplified picture of the three prognostic variables related to soil moisture: *catchment deficit* (CATDEF), *root zone excess* (RZEXC), and *surface excess* (SRFEXC). In effect, the vertical profile of soil moisture at each point in each computational unit (related to CATDEF; see Figure 5) is determined by the equilibrium soil moisture profile from the surface to the (spatially varying) water table (defined by a balance of gravity and capillary forces) and by two additional variables that describe deviations from the equilibrium profile: the average deviation in a 1 m root zone layer (RZEXC), and the average deviation in a 2 cm surface layer (in the GMAO operational configuration; SRFEXC). For SMAP, the surface layer depth will be changed to 5 cm to correspond more closely to the instrument's observing depth.

Variable name	**Description**	**Units**
TC1	Surface temperature (saturated area)	K
TC2	Surface temperature (unsaturated area)	K
TC4	Surface temperature (wilting area)	K
QA1	Canopy air specific humidity (saturated area)	$kg\ kg^{-1}$
QA2	Canopy air specific humidity (unsaturated area)	$kg\ kg^{-1}$
QA4	Canopy air specific humidity (wilting area)	$kg\ kg^{-1}$
CAPAC	Interception reservoir	$kg\ m^{-2}$
CATDEF	Catchment deficit	$kg\ m^{-2}$
RZEXC	Root zone excess	$kg\ m^{-2}$
SRFEXC	Surface excess	$kg\ m^{-2}$
GHT1, GHT2, …, GHT6	Ground heat content (layers 1-6)	$J\ m^{-2}$
WESN1, WESN2, WESN3	Snow water equivalent (layers 1-3)	$kg\ m^{-2}$
SNDZ1, SNDZ2, SNDZ3	Snow depth (layers 1-3)	m
HTSN1, HTSN2, HTSN3	Snow heat content (layers 1-3)	$J\ m^{-2}$

Table 3. Prognostic variables of the Catchment model.

As indicated in Figure 5, the Catchment model differs from traditional, layer-based models by including an explicit treatment of the spatial variation of soil water and water table depth within each computational unit (that is, within each 9 km EASE grid cell for L4_SM) based on the statistics of the catchment topography. This spatial variation enters into the calculation of moisture diffusion between the root zone and lower soil moisture storage. Extensive preprocessing produces a pre-computed functional relationship between RZEXC, CATDEF, and the amount of moisture transferred between the two in a given time step, a functional relationship that is based on a spatially distributed set of one dimensional Richard's equation calculations, each representing moisture transport at some location in the catchment and each performed on a soil column fitted with high vertical resolution. The transfer of moisture between the 0-5 cm surface layer and the root zone, of particular relevance to SMAP, is computed similarly, though without a spatially distributed component; a highly-resolved, one-dimensional representation of the root zone is used to pre-compute a functional relationship between the moisture variables and the amount of moisture transferred between SRFEXC and RZEXC within the time step.

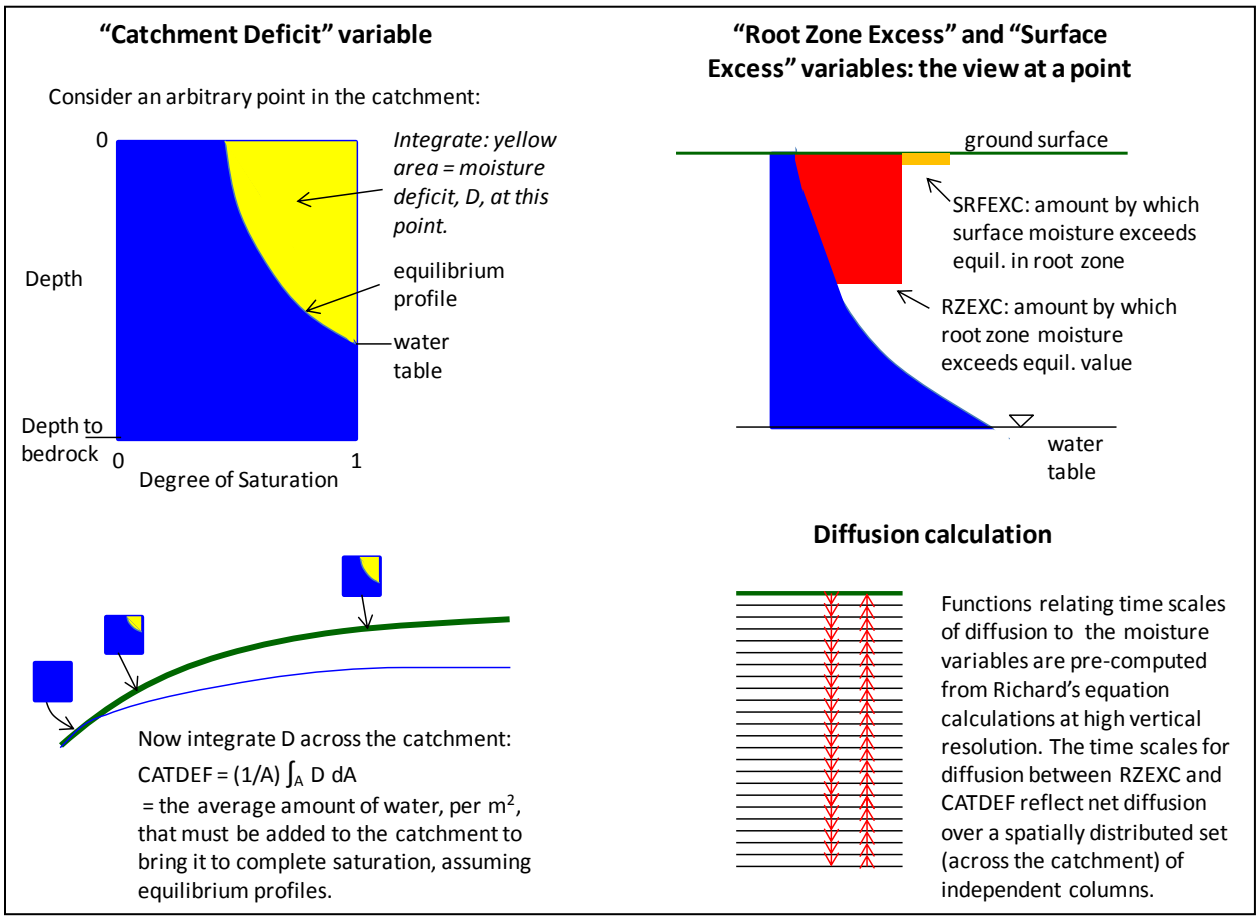

Figure 5. Unique elements of the Catchment land surface model related to the diffusion of moisture between the 0-5 cm surface zone and the remainder of the soil profile. Shown are descriptions of the three moisture prognostic variables (CATDEF, RZEXC, and SRFEXC; Table 3) and an indication of how the transfer of moisture between the variables is computed.

The treatment of spatial heterogeneity also allows the diagnostic separation of the catchment into saturated, unsaturated, and wilting sub-grid areas. The sizes of these sub-grid areas vary dynamically; wetter conditions, for example, expand the saturated sub-grid area and reduce the wilting sub-grid area. The surface energy balance is computed separately for each sub-grid area using physics specific to the corresponding hydrological regime. This entails the monitoring of independent prognostic temperature variables for each sub-grid area (TC1, TC2, and TC4). The three surface temperature prognostic variables interact with an underlying heat diffusion model for soil temperature (consisting of six layers with depths equal to about 0.1, 0.2, 0.4, 0.75, 1.5, 10 m from top to bottom) that is common to all three sub-grid areas. Surface runoff processes are computed separately for each sub-grid area, again using hydrological regime-specific physics, whereas subsurface baseflow is computed directly from the diagnosed spatial distribution of water table depth. A snow model component describes the state of snow pack in terms of snow water equivalent (WESN), snow depth (SNDZ), and snow heat content (HTSN; three layers for each variable). The time step for the model integration is 20 minutes. Table 7a includes a list of the most important Catchment model parameters.

The salient feature of the land model integration is that it uses meteorological forcing inputs that rely on observed data as much as possible. *Reichle et al*. [2011], *Yi et al*. [2011], and *Holmes et al*. [2011] provide a comprehensive assessment of large-scale land surface estimates derived with the Catchment model as part of the MERRA reanalysis and demonstrate that the Catchment model is a state-of-the-art global land surface model.

4.1.2b The microwave radiative transfer model

The Catchment model has been augmented with a microwave radiative transfer model that transforms the simulated surface soil moisture and temperature fields into model estimates of L-band brightness temperature (at the 9 km scale). Like the L2_SM_P and L2_SM_AP algorithms, the L4_SM algorithm uses the "tau-omega" model, an approximation of the radiative transfer processes that is appropriate for low frequency microwave emission. In this model, "tau" is the vegetation optical depth and "omega" is the single-scattering albedo. A layer of vegetation over the soil attenuates the emission from the soil and adds to the total radiative flux with its own emission. Assuming that scattering within the vegetation is negligible at L-band frequencies, the vegetation may be treated mainly as an absorbing layer for the soil moisture signal (see L2_SM_P ATBD for details; section 8).

The parameterizations of the microwave model represent a tradeoff between the need to adequately represent the key effects of surface characteristics on microwave signatures at the spatial scale of interest, and the need for a sufficiently simple representation for application to satellite retrieval algorithms. The microwave model incorporates the effects of dynamic features (including surface soil moisture and soil temperature), and static or slowly-varying features such as soil texture, soil surface roughness, land-cover and vegetation type, and vegetation water content. Effects of atmospheric variability are assumed negligible at L-band for non-raining conditions. Table 7a includes a list of the most important ancillary parameters used in the microwave radiative transfer model.

In terms of the choices for roughness and vegetation parameterizations and the corresponding ancillary parameter datasets, the configuration of the L4_SM microwave model will correspond as closely as possible to that of the L2_SM_P and L2_SM_AP algorithms. However, there are important reasons why the configuration of the microwave model for L4_SM will likely differ from that used for the Level 2 algorithms (section 4.1.3).

4.1.2c The ensemble Kalman filter

The L4_SM algorithm is built on the ensemble Kalman filter (EnKF) – a Monte-Carlo variant of the Kalman filter [*Evensen*, 2003]. The idea behind the EnKF is that a small ensemble of model trajectories captures the relevant parts of the error structure. Each member of the ensemble experiences perturbed instances of the observed forcing fields (representing errors in the forcing data) and/or randomly generated noise that is added to the model parameters and prognostic variables (representing errors in model physics and parameters). The error covariance matrices that are required for the filter update can then be diagnosed from the spread of the ensemble at the update time. The EnKF is flexible in its treatment of errors in model dynamics and parameters. It is also very suitable for modestly nonlinear problems and has become a popular choice for land data assimilation [*Andreadis and Lettenmaier*, 2005; *Durand and Margulis*, 2007; *Kumar et al.*, 2008a,b; *Pan and Wood,* 2006; *Reichle et al.*, 2002a,b; *Zhou et al.*, 2006].

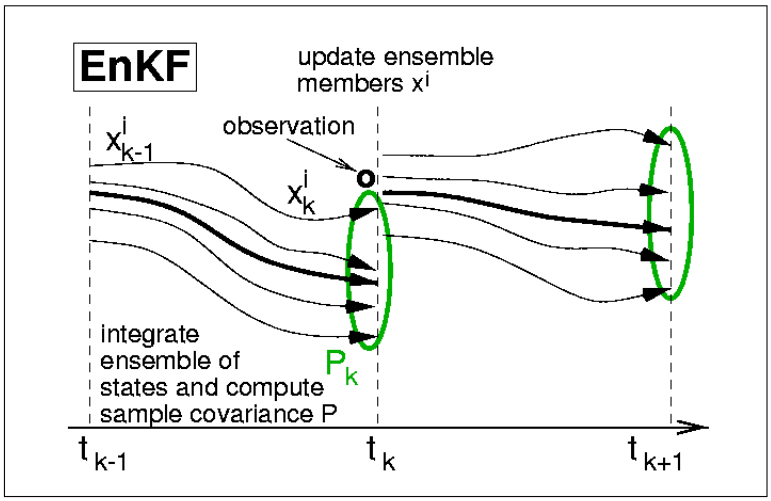

Figure 6. The ensemble Kalman filter (EnKF).

The EnKF works sequentially by performing in turn a model forecast and a filter update (Figure 6). Formally, the forecast step for ensemble member i can be written as

$$x_{t,i}^{-} = f(x_{t-1,i}^{+}, q_{t,i}) \quad (1)$$

where $x_{t,i}^{-}$ and $x_{t-1,i}^{+}$ are the forecast (denoted with $^{-}$) and analysis (denoted with $^{+}$) state vectors at times t and t-1, respectively, of the i-th ensemble member. The model error (or perturbation

vector) is denoted with $q_{t,i}$ and its covariance with Q_t. Each ensemble member represents a particular realization of the possible model trajectories with certain random errors in model parameters and/or a particular set of errors in forcing.

With the observations available at time t, the state of each ensemble member is updated to a new value. First, the filter update produces increments at time t that can be written as

$$\Delta x_{t,i} = K_t (y_{t,i} - H_t x_{t,i}^-) \quad (2)$$

where $y_{t,i}$ denotes the observation vector (suitably perturbed) and H_t is the observation operator (which is written as if it was linear for ease of notation, but in practice the update is solved without explicitly computing H_t, [*Keppenne*, 2000]). Next, the analyzed state vector is obtained as $x_{t,i}^+ = x_{t,i}^- + \Delta x_{t,i}$. The Kalman gain matrix K_t is given by

$$K_t = P_t H_t^T (H_t P_t H_t^T + R_t)^{-1} \quad (3)$$

where P_t is the forecast error covariance (diagnosed from the ensemble $x_{t,i}^-$), R_t is the observation error covariance, and superscript T denotes the matrix transpose. Simply put, the Kalman gain K_t represents the relative weights given to the model forecast and the observations based on their respective uncertainties, along with the error correlations between different elements of the state vector. If the system is linear, if its model and observation error characteristics satisfy certain assumptions (including Gaussian, white, and uncorrelated noise), and if the input error parameters are correctly specified, the Kalman gain of equation (3) is optimal in the sense of minimum estimation error variance. In other words, the updated state is mathematically the best estimate of the state possible given the observations, the model prediction, and the estimated errors of both. The reduction of the uncertainty resulting from the update is reflected in the reduction of the ensemble spread. Note that the ensemble of model trajectories in the EnKF naturally yields error estimates for the assimilation products.

4.1.2d Freeze-thaw analysis

The assimilation of SMAP freeze/thaw state (L3_FT_A) observations is conceptually similar to the assimilation of snow cover observations. In both cases, the observed variable is, at least at a finer spatial scale, essentially a binary measurement. Generally, binary measurements cannot be assimilated with the EnKF, because the EnKF requires continuous variables (such as water or heat reservoirs). This restriction can be circumvented, however, for snow cover observations because in land models, fractional snow cover for a given model grid cell can be related to a continuous prognostic variable (such as SWE) via a snow depletion curve. By aggregating high-resolution measurements into fractional snow cover observations (at the scale of the land model) the EnKF could still be used [*De Lannoy et al.,* 2010; *De Lannoy et al.,* 2011]. The same does not apply to freeze-thaw observations, because there is no equivalent to the snow depletion curve for the land model's freeze-thaw state. Consequently, for the assimilation of freeze-thaw observations in the baseline L4_SM algorithm we will adapt the rule-based (non-EnKF) approaches that have been developed to assimilate snow cover observations [*Rodell and Houser*, 2004; *Zaitchik and Rodell*, 2009].

Because the radar and radiometer measurements are not informative of soil moisture under frozen soil conditions, a given 9 km grid cell is never included simultaneously in the freeze-thaw analysis and the soil moisture analysis (Figure 4). In particular, a given 9 km grid cell will be included in the soil moisture analysis (section 4.1.2e) only if both the observations and the model indicate thawed conditions. If the observations and the model agree on frozen conditions, there will be no further analysis step for the 9 km grid cell in question. If the model forecast and the corresponding SMAP observations disagree on the freeze-thaw state, that is, if the model indicates frozen conditions and the observation indicates thawed conditions (or vice versa), the model prognostic variables (Table 3) will be adjusted to match the model's freeze-thaw state to the observation in a freeze-thaw analysis. Adjustments will primarily be made to the forecast surface soil temperatures (TC1, TC2, TC4; Table 3), the soil heat content (GHT1-6; Table 3), and, if snow is present, the snow prognostic variables (WESN1-3, SNDZ1-3, and HTSN1-3; Table 3).

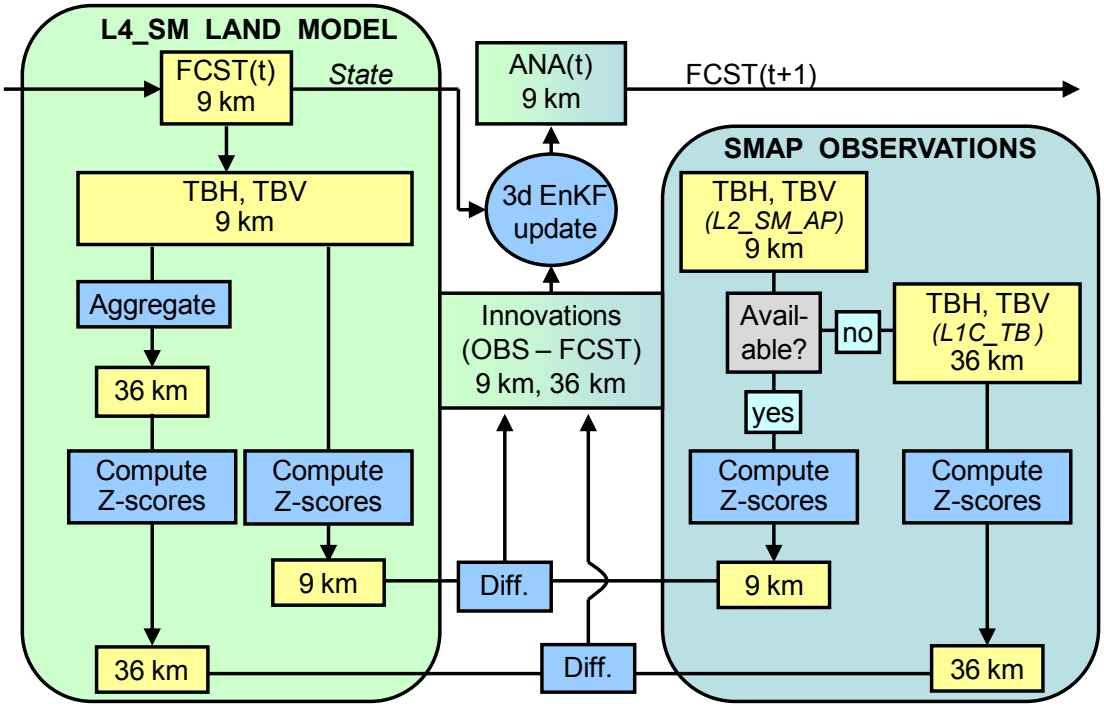

Figure 7. The L4_SM soil moisture analysis.

4.1.2e Soil moisture analysis

Figure 7 summarizes the soil moisture analysis of the L4_SM algorithm. The **state vector** x for the soil moisture analysis of the **baseline** algorithm consists of seven Catchment model prognostic variables (*catchment deficit*, *root zone excess*, *surface excess*, three *surface temperature prognostic* variables, one each for the saturated, unsaturated, and wilting sub-grid

areas, and the first-layer *ground heat content*; Table 3 and section 4.1.2a) at each computational element (9 km grid cell) that is included in the soil moisture analysis (see also Figure 4). Formally, the forecast state vector for the soil moisture analysis is

$$\mathbf{x}^- = \begin{bmatrix} \mathbf{x}_1^- \\ \mathbf{x}_2^- \\ \cdots \\ \mathbf{x}_{N9}^- \end{bmatrix}, \qquad \text{where} \quad \mathbf{x}_j^- = \begin{bmatrix} \text{SRFEXC\_FCST}_j \\ \text{RZEXC\_FCST}_j \\ \text{CATDEF\_FCST}_j \\ \text{TC1\_FCST}_j \\ \text{TC2\_FCST}_j \\ \text{TC4\_FCST}_j \\ \text{GHT1\_FCST}_j \end{bmatrix}, \qquad (4)$$

N9 is the number of 9 km grid cells included in the soil moisture analysis, and j=1...N9. For clarity we omitted the subscripts for time and ensemble member[3]. Since the ancillary data for the vegetation opacity τ are relatively uncertain, augmentation of the state vector with τ will be considered as an **optional** variant. In this case, the dynamical model for τ will be a relaxation to the ancillary value, subject to intermittent forcing by the increments from the EnKF update.

As mentioned in section 4.1.1, the L4_SM baseline algorithm assimilates downscaled (9 km) brightness temperatures in H- and V-polarization (TBH09_OBS and TBV09_OBS) when and where available from the L2_SM_AP product. However, high-resolution backscatter data are not always available to generate the downscaled brightness temperatures. For example, during afternoon overpasses high-resolution radar data are collected only north of 45° N because of resource limitations. Even if high-resolution backscatter data are available, the L2_SM_AP algorithm may not always provide downscaled brightness temperatures. If, for a given time and location, downscaled (9 km) brightness temperatures are not available, the 36 km brightness temperature values from the L1C_TB product in H- and V-polarization (TBH36_OBS and TBV36_OBS) will be assimilated. Note that we will *not* assimilate the 36 km brightness temperatures for a given time and location if downscaled (9 km) values are available for that time and location.

Recently, the L-band brightness temperatures generated by the Catchment model and its associated microwave radiative transfer model described above have been calibrated (separately for each location) to match the climatology of SMOS observations (section 4.1.3). While the model calibration yields largely unbiased modeled brightness temperatures (with respect to SMOS), residual model biases remain and are primarily related to seasonal variations in bias. Moreover, it is not clear to what extent the SMOS observations are impacted by low-level RFI and may themselves be biased. These unavoidable biases in the model and the observations must be addressed as part of the data assimilation system (section 2.2). Specifically, the brightness

[3]For reasons of numerical accuracy in the subsequent matrix calculations, the Catchment model prognostic variables for each element of the state vector are multiplied with appropriate constant scaling (or unit conversion) factors to ensure that all elements of the state vector are approximately of the same order of magnitude. This scaling should not be confused with the a posteriori climatological scaling of surface soil moisture estimates from the climatology of the Catchment model into that of the L2_SM_AP product.

temperatures will be converted to **Z-scores** (or standard normal deviates; separately for H- and V-polarization) before the innovations are computed (Figure 7). Since brightness temperature is strongly impacted by surface temperature, it is important to resolve the seasonal and diurnal cycles of the climatology, which limits us to using first and second moments (as opposed to matching cumulative distribution functions (cdfs)) [*Reichle et al.,* 2010].

Specifically, the **observed** brightness temperatures for a given time and location (corresponding subscripts omitted for clarity) are converted into non-dimensional standard normal deviates (or Z-scores) through:

$$\text{TBprr\_OBS\_ZSCORE} = \frac{(\text{TBprr\_OBS} - \text{TBprr\_OBS\_MEAN})}{\text{TBprr\_OBS\_STD}} \tag{5a}$$

where p=H or p=V indicates the polarization, rr=09 or rr=36 indicates the resolution (in km), TBprr_OBS_MEAN is the *multi-year* mean value of the observed brightness temperature (for a given location, day-of-year, and overpass time-of-day), and TBprr_OBS_STD is the corresponding value for the standard deviation (Table 7b). SMOS brightness temperatures will provide a useful early estimate of the SMAP brightness temperature climatologies. The brightness temperatures produced by the land model are calibrated to the SMOS climatology prior to the launch of SMAP (section 4.1.3). The SMOS climatology will also be used initially in L4_SM production until sufficient SMAP observations have been accumulated. Thereafter, a SMAP-only climatology will be used for recalibrating the land model and for generating and reprocessing the L4_SM product.

The **model** predictions of brightness temperature are similarly converted into Z-scores:

$$\text{TBprr\_FCST\_ZSCORE} = \frac{(\text{TBprr\_FCST} - \text{TBprr\_MOD\_MEAN})}{\text{TBprr\_MOD\_STD}} \tag{5b}$$

where for clarity we again omitted the subscripts for time, location, and ensemble member.

Following equation (2), the innovations vector $(y - H\bar{x})$ will thus be computed by differencing the H- and V-polarization brightness temperature *Z-scores* from the observations and the Catchment model forecast. If downscaled (9 km) brightness temperature are available for a given 36 km grid cell, up to $2\cdot(36/9)^2=32$ elements from that grid cell are included in the innovations vector. Otherwise, the 36 km grid cell in question only contributes two elements to the innovations vector. Formally, the **observation vector** is therefore

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{N36} \end{bmatrix}, \tag{6}$$

$$\text{where} \quad y_k = \begin{cases} \begin{bmatrix} \text{TBH09\_OBS\_ZSCORE}_1 \\ \text{TBV09\_OBS\_ZSCORE}_1 \\ \text{TBH09\_OBS\_ZSCORE}_2 \\ \text{TBV09\_OBS\_ZSCORE}_2 \\ \ldots \\ \text{TBH09\_OBS\_ZSCORE}_{M(k)} \\ \text{TBV09\_OBS\_ZSCORE}_{M(k)} \end{bmatrix}_k & \text{if 9 km TB available from L2\_SM\_AP} \\ \begin{bmatrix} \text{TBH36\_OBS\_ZSCORE} \\ \text{TBV36\_OBS\_ZSCORE} \end{bmatrix}_k & \text{otherwise.} \end{cases}$$

Here, N36 is the number of 36 km grid cells with brightness temperature observations in the soil moisture analysis, k=1...N36, and M(k) is the number of assimilated downscaled brightness temperature observations within the k-th 36 km grid cell (if available). Again, we omitted the time subscript for clarity. As mentioned above, the observations vector is suitably perturbed before assimilation (equation (2)).

The corresponding vector of **model predictions** of the 9 km and 36 km brightness temperature Z-scores are computed from the Catchment model forecast, that is,

$$\mathrm{H}\mathbf{x}^- = \begin{bmatrix} \mathrm{H}\mathbf{x}_1^- \\ \mathrm{H}\mathbf{x}_2^- \\ \ldots \\ \mathrm{H}\mathbf{x}_{N36}^- \end{bmatrix}, \tag{7}$$

$$\text{where} \quad Hx_k^- = \begin{cases} \begin{bmatrix} \text{TBH09\_FCST\_ZSCORE}_1 \\ \text{TBV09\_FCST\_ZSCORE}_1 \\ \text{TBH09\_FCST\_ZSCORE}_2 \\ \text{TBV09\_FCST\_ZSCORE}_2 \\ \ldots \\ \text{TBH09\_FCST\_ZSCORE}_{M(k)} \\ \text{TBV09\_FCST\_ZSCORE}_{M(k)} \end{bmatrix}_k & \text{if 9 km TB available from L2\_SM\_AP} \\ \begin{bmatrix} \text{TBH36\_FCST\_ZSCORE} \\ \text{TBV36\_FCST\_ZSCORE} \end{bmatrix}_k & \text{otherwise.} \end{cases}$$

Again, equations (6) and (7) define the innovations vector needed for the analysis update (equation (2)). Note that the observation operator defined in equation (7) contains all the processing steps required to map the state vector x (equation (4)) into a model prediction of the observed values that can then be directly differenced with the observation vector y (equation (6)). The observation operator thus includes (i) the transformation of soil moisture and soil temperature fields into brightness temperatures via the microwave radiative transfer model at 9 km resolution, (ii) the aggregation from 9 km to 36 km (only!) for locations where 36 km brightness temperature observations from L1C_TB are assimilated, and (iii) the Z-score calculation of equation (5b).

Finally, the increments are computed in units of Catchment model prognostic variables following [*Keppenne,* 2000] (see discussion of equations (2) and (3)). In the baseline L4_SM system, the EnKF will be implemented with three-dimensional ("3d") updates [*Reichle and Koster,* 2003], that is, the increments for a given 9 km grid cell are affected by all observations within a certain radius of influence, and not just by the observations that cover the grid cell in question. The computationally simpler one-dimensional ("1d") algorithm that uses only local observations will be considered as a down-scope option if computing resources should unexpectedly prove insufficient for the 3d implementation. The radius of influence for the "3d" algorithm is determined by the spatial error correlation scales and is expected to be no more than a few hundred kilometers [*Reichle and Koster,* 2003]. It has been shown that the ensemble filter works adequately with 12 ensemble members [*Reichle et al.*, 2007; *Liu et al.*, 2011]. To reduce sampling errors, we consider using at least 24 ensemble members (resources permitting, section 4.2).

The primary outputs of the L4_SM algorithm are surface and root zone soil moisture in percentile units (section 2.3.2). Furthermore, L4_SM surface soil moisture estimates in

volumetric units (m^3 m^{-3}) that are consistent with the L2_SM_AP surface soil moisture climatology (SFMC_L2CLIM in Table 6a) will be generated via the cdf matching approach introduced by *Reichle and Koster* [2004] (section 2.2). By construction, the cdf matching approach requires a time series record that is long enough such that its climatology can be estimated reliably. To use the cdf matching approach with new satellite observations such as from SMAP, the scaling parameters can be estimated through the ergodic substitution of variability in space for variability in time. Using historic soil moisture satellite retrievals from SMMR, *Reichle and Koster* [2004] were able to show that soil moisture biases can be adequately addressed with a one-year satellite record (Figure 8). A preliminary estimate of the climatology of the L2_SM_AP soil moisture retrievals will be obtained by using SMOS brightness temperature observations to derive surface soil moisture with the SMAP L2_SM_P algorithm (which is an integral part of the L2_SM_AP algorithm). This preliminary climatology will be used in the L4_SM algorithm to generate the SFMC_L2CLIM output until enough SMAP L2_SM_AP retrievals have been generated to use a SMAP-only climatology. Note that the soil moisture climatology of the Catchment model will be available prior to launch. Following *Drusch et al.* [2005], we will use a third-order polynomial approximation to parameterize the cdf. The cdf parameters are provided as part of the L4_SM "clim" output file (Table 7b).

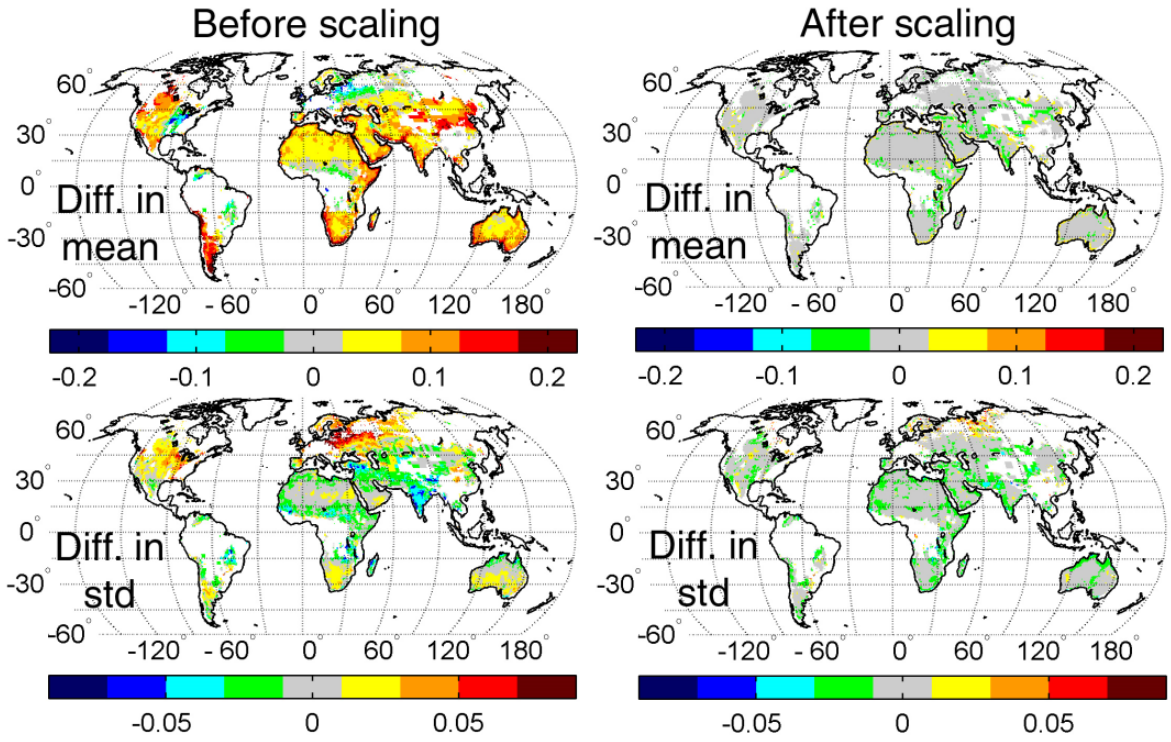

Figure 8. Difference in the 1979-1987 (top) mean and (bottom) standard deviation of satellite and model soil moisture (left) before and (right) after scaling. Units are volumetric soil moisture in m^3 m^{-3} (absolute soil moisture typically ranges between 0 and 0.5 m^3 m^{-3}). Scaling of the satellite soil moisture is based only on data from a single year (1979) using the ergodic substitution of variability in space for variability in time. From [*Reichle and Koster*, 2004].

4.1.2f Algorithm upscope options

We consider three algorithm upscope options. The first option is to implement an ensemble smoother instead of the ensemble filter. A smoother uses "future" observations and is particularly attractive for reprocessing. The second optional extension is a dynamic bias estimation and correction algorithm that may supersede or work together with the assimilation of brightness temperature Z-scores (standardized anomalies; section 4.1.2). The third optional extension is an adaptive filtering module that dynamically estimates key input error parameters. All three upscope options have the potential to improve the L4_SM product. The second and third optional extension algorithms are already implemented in the NASA GEOS-5 land data assimilation system. More development, however, is needed to implement the optional algorithms in such a way that they can work together. Unlike the baseline algorithm, which has already been tested with SMAP precursor observations, the optional extensions have yet to be thoroughly tested with satellite data and will be used only if warranted by the test results.

Upscope option 1: Smoothing

The first upscope option is the implementation of a smoother (or batch) algorithm. In a smoother, measurements at different times within an assimilation (time) interval are processed simultaneously [*Dunne and Entekhabi*, 2006; *Reichle,* 2008]. The state estimate at the initial time is therefore based on "future" observations from the entire assimilation interval. For most applications, such a smoothing solution can only cover a relatively short time interval – typically just 12 h in data assimilation systems used for weather prediction.

A smoother has obvious advantages. The intermittent character of rain storms and subsequent dry-down periods can be captured more accurately. In practice, however, there are important obstacles. The first consideration is the increased computational cost. For a given amount of resources, the benefit of a smoothing algorithm must be traded off, for example, against using more ensemble members and implementing the data assimilation system at the relatively high resolution of 9 km globally. Relatively cheap implementations of a smoother exist, but these come with the cost of providing flux estimates that are identical to those obtained from filtering algorithms, thereby negating what may be the chief benefit of the smoother. Moreover, the intermittent character of precipitation forcing limits the information that "future" observations can provide soil moisture conditions prior to the observation time.

We will assess the feasibility and advantages of a smoothing algorithm for the L4_SM product as part of the algorithm development to the extent possible given the available resources.

Upscope option 2: Dynamic bias estimation

An alternative strategy to alleviate satellite-model biases is to dynamically estimate bias parameters along the lines of the algorithm developed by *Dee* [2005]. This approach was used for soil moisture by *De Lannoy et al.* [2007] and by *Reichle et al.* [2010] for land surface temperature. This dynamic bias estimation approach is based on a second Kalman filter for bias estimation (in addition to the Kalman filter for state estimation described above). Assume that

we have a bias estimate b_{t-1}^{+} at time t-1. Furthermore, assume that this bias estimate can be propagated to time t with a simple bias evolution model

$$b_t^{-} = \eta\, b_{t-1}^{+} \tag{8}$$

that relaxes the bias estimates to zero ($\eta<1$). The use of a relaxation factor is a prudent strategy for experiments that cover many seasons. Because observations may not be available for extended periods, relaxing the bias estimate to zero is safer than keeping the latest bias estimate through seasons for which it may not be appropriate.

Next, we compute a bias-corrected model forecast

$$\xi_{t,i}^{-} = x_{t,i}^{-} - b_t^{-} \tag{9}$$

that is used in the state update equation (2) (instead of the biased model forecast $x_{t,i}^{-}$). From ensemble average innovations (computed as $y_t - H_t\, \xi_t^{-} \equiv E\{y_{t,i} - H_t\, \xi_{t,i}^{-}\}$, where $E\{\cdot\}$ is the ensemble mean operator), we can then update the bias via

$$b_t^{+} = b_t^{-} - \lambda\, K_{x,t}\, (y_t - H_t\, \xi_t^{-}). \tag{10}$$

A key assumption of this algorithm is that the bias error covariance $P_{b,t}$ is a small fraction of the state error covariance, that is $P_{b,t} = \lambda\, P_{x,t}$, which implies that the gain for the bias ($K_{b,t}$) can be computed as a fraction (say, $\lambda = 0.2$) of the gain for the state, which has already been computed.

Equation (10) implies that in practice, the bias estimates can be thought of as a moving window time average of the soil moisture increments. Unlike the a priori scaling (or Z-score assimilation) approach, the dynamic bias estimation strategy adapts to slow changes in bias over time. A disadvantage of this strategy is the assumption that only the land model soil moisture but not the satellite soil moisture is biased. The dynamic bias estimation and correction algorithm may supersede or work together with the assimilation of brightness temperature Z-scores. *Reichle et al.* [2010] found that for land surface temperature the best strategy depended on the details of the land model formulation.

Upscope option 3: Adaptive filtering

The input error parameters that are required by the data assimilation system (section 4.1.3) are a source of uncertainty. Errors in their specification may be alleviated with an adaptive filtering approach [*Reichle et al.*, 2008]. The central idea behind adaptive filtering methods is that internal diagnostics of the assimilation system should be consistent with the values that are expected from input error parameters provided to the data assimilation system. The most commonly used diagnostics for adaptive filtering are based on the observation-minus-forecast residuals or innovations (computed in the EnKF as $v_t \equiv E\{y_{t,i} - H_t\, x_{t,i}^{-}\}$, where $E\{\cdot\}$ is the ensemble mean operator). For a linear system operating under optimal conditions, the lagged innovations covariance is

$$E[v_t\, v_{t-k}^T] = \delta_{k,0}\,(H_t\, P_t\, H_t^T + R_t) \tag{11}$$

where $E[\cdot]$ is the expectation operator and $\delta_{k,0}$ is the Kronecker delta. Equation (11) implies that the innovations sequence is uncorrelated in time and that its covariance is equal to the sum of the forecast error covariance $H_t\, P_t\, H_t^T$ (in observation space) and the observation error covariance R_t. Now recall that the forecast error covariance P depends on the model error covariance Q. If the innovations show less spread than expected, the input error covariances (Q and/or R) are too large, and if the innovations show more spread than expected, the input error covariances are too small. Such information can be used for adaptive tuning of Q and/or R.

Alternative diagnostics are based on the analysis departures $w_t \equiv E\{y_{t,i} - H_t\, x_{t,i}^+\}$ and the (observation space) analysis increments $u_t \equiv E\{H_t(x_{t,i}^+ - x_{t,i}^-)\}$. For linear systems operating under optimal conditions we have [*Desroziers* et al., 2005]

$$E[u_t\, v_t^T] = H_t\, P_t\, H_t^T \tag{12}$$

$$E[w_t\, v_t^T] = R_t \tag{13}$$

Equations (12) and (13) suggest a simple way of estimating the model and observation error covariances separately. A flow chart of the adaptive EnKF algorithm is provided in Figure 9. After appropriate initialization, the adaptive EnKF starts with a regular EnKF forecast and update step (equations (1)-(3)), along with the computation of the innovations, the analysis departures, and the analysis increments. The adaptive module follows thereafter. In our formulation, the adaptive filter is estimating the ratios of the true input error variances and their *initial* values, that is we are ultimately trying to determine – separately for each location – the scalar factors $\alpha_Q = \text{trace}(Q_{true}) / \text{trace}(Q_{t=0})$ and $\alpha_R = \text{trace}(R_{true}) / \text{trace}(R_{t=0})$, which are also referred to as "adaptive tuning factors." To this end, the adaptive module starts with estimating the terms in equations (12) and (13). Because of the substantial heterogeneity of the land surface, we compute the desired statistic from the relevant time series for each catchment individually (and thereby reduce equations (12) and (13) to sets of scalar equations). The most recent estimate of $E[u\, v^T]$, for instance, is approximated as an exponential moving average, that is, at update time t we approximate $E[u\, v^T] \approx MA[u\, v^T]_t$, where the exponential moving average is denoted with $MA[\cdot]$ and defined as $MA[u\, v^T]_t \equiv (1 - \gamma)\, MA[u\, v^T]_{t-1} + \gamma\, u_t\, v_t^T$, with an ad-hoc choice of γ=0.02 for the experiments presented here. Similarly, we compute time-filtered estimates of $E[w\, v^T]$, $H\, P\, H^T$, and R (Figure 9a).

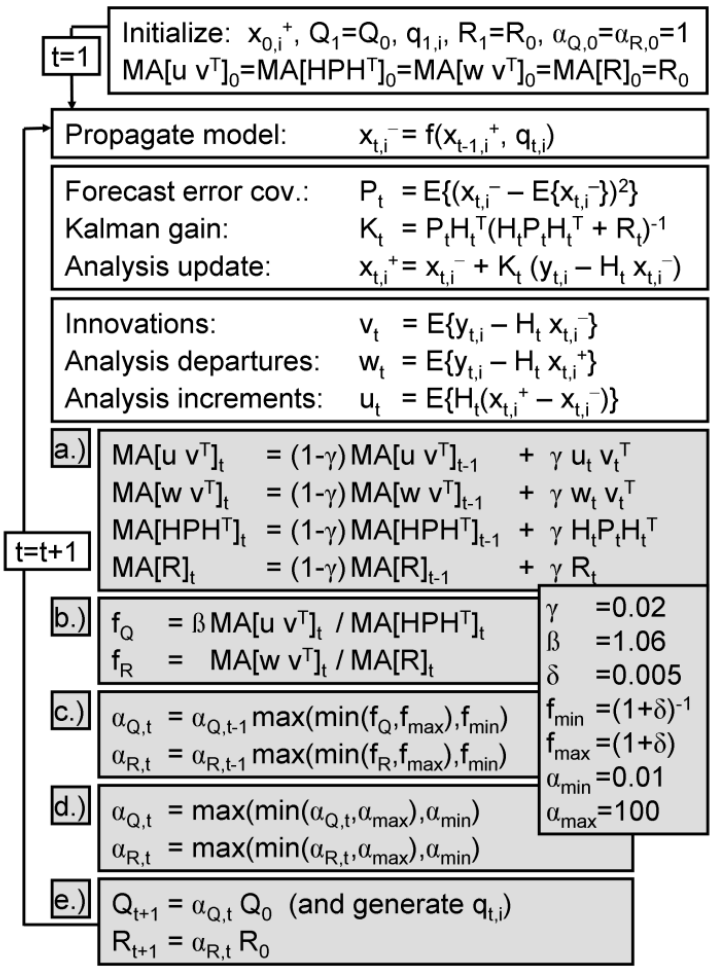

Figure 9. Adaptive EnKF algorithm. Initial guesses for model and observation error covariances are denoted with Q_0 and R_0, respectively. Shaded boxes indicate the adaptive module of the EnKF. "MA[·]" denotes exponential moving average, E{·} denotes ensemble mean. See text for symbols and further discussion. From [*Reichle et al.*, 2008b].

Next, we determine the ratio of actual covariances and expected covariances, that is, we compute the ratios $f_Q = \beta\ MA[u\ v^T]_t / MA[H\ P\ H^T]_t$ and $f_R = MA[w\ v^T]_t / MA[R]_t$ (Figure 9b; the factor β will be explained below). We then use these ratios to update the adaptive tuning factors α_Q and α_R based on the idea (expressed in equations (12) and (13)) that these ratios should be close to unity for correct input error parameters. Because the moving average estimates of the terms in equations (12) and (13) are very noisy even after temporal smoothing, it is necessary to further restrict the ratios f_Q and f_R to the interval $[(1+\delta)^{-1}, (1+\delta)]$, with δ=0.005 used here (Figure 9c). In practice, this means that α_Q and α_R are decreased or increased by only a very small fraction at each update time, no matter how far the original ratios f_Q and f_R are from unity. While slowing down convergence considerably, such a restriction is necessary for stabilizing the algorithm. In addition, we restrict α_Q and α_R to the interval [0.01, 100] (Figure 9d). Finally, we derive the latest estimates of the true Q and R by scaling the original input model parameters $Q_{t=0}$ and $R_{t=0}$ with the latest estimates of α_Q and α_R, respectively (Figure 9e). Note that for

lognormally distributed, multiplicative perturbations (such as are applied to precipitation and shortwave radiation forcing, section 4.1.3), the covariance scaling is applied to the corresponding normal deviates. The "scaled" lognormal perturbations are computed by transforming the scaled normal deviates into lognormal space. Note also that multiplicative perturbations (as opposed to additive perturbations) exacerbate the noisy character of the ratios f_Q and f_R.

It is important to note that in an EnKF-based land assimilation system we are interested in scaling the model error covariance Q, and not the forecast error covariance P (as in [*Dee*, 1995] and [*Desroziers* et al., 2005]). Our adaptive approach therefore makes a linearity assumption between P and Q that is only valid approximately. In particular, we found that the ratio f_Q (Figure 9b) is systematically underestimated for perfect input error parameters unless an empirical factor β is introduced. This factor can be determined in a twin experiment that uses perfect input error parameters. If the factor β is omitted, the adaptive algorithm tends to produce slightly worse results (when compared to the non-adaptive algorithm) in the case of correct initial input error parameters. It is, however, highly unlikely that the initial guess of input error parameters is already perfect, and the adaptive filter does yield significant improvements for most choices of input error parameters even when the factor β is omitted. It must also be noted that theoretically, the adaptive algorithm relies on stationary conditions, and Q_{true} and R_{true} should not vary with time. Nevertheless, slow variations of Q_{true} and R_{true} on seasonal time scales and longer can be handled by the algorithm without significant loss of performance.

4.1.3 Ancillary Data Requirements

Aside from SMAP observations, the data assimilation system requires initialization, parameter and forcing inputs for the Catchment land surface model, as well as input error parameters for the ensemble-based data assimilation system.

4.1.3a Catchment land surface model parameters

The Catchment model requires topography, soil, and vegetation data at all computational elements in the chosen spatial discretization. A full set of these Catchment model parameters is available as part of the GEOS-5 modeling system [*Rienecker et al*., 2008]. To the extent possible, Catchment model parameters will be adjusted for consistency with land surface parameters that are used by other SMAP products. The current version of the Catchment model parameters related to subsurface moisture has been derived from 1 km global topographic data (digital elevation model). A project is underway to update these parameters using higher resolution (90 m) topographic information. We expect that the updated model parameters will be available prior to the launch of SMAP and can thus be used to generate the L4_SM product. Catchment model ancillary parameter inputs are provided as part of the L4_SM data product (Table 7a).

4.1.3b Microwave radiative transfer model configuration and parameters

A variety of parameterizations exists for soil dielectric mixing, soil roughness effects, and vegetation opacity. Recent publications by *de Rosnay et al*. [2009], *Drusch et al*. [2009], and

Sabater et al. [2011] assess the performance of these parameterizations in the context of global modeling. The current configuration of the microwave radiative transfer model for the L4_SM algorithm includes the [*Wang and Schmugge*, 1980] soil dielectric mixing model, the [*Wang and Choudhury*, 1981] soil roughness model, the [*Jackson and Schmugge*, 1991] vegetation opacity model, and the [*Pellarin et al*., 2003] atmospheric correction. See Table 7a in section 6 for a list of the most important ancillary parameters that are required for this configuration. Note that some of the soil parameters of the Catchment model may differ from those of the microwave radiative transfer model because the two models describe soil profiles of different depths. Similarly, the vegetation class inputs for the two models may differ because the two models may be associated with different vegetation classifications.

Eventually, the configuration and parameters of the L4_SM microwave model will correspond as closely as possible to that of the L2_SM_P and L2_SM_AP algorithms. However, there are important reasons why the model configuration and parameters may differ between the L4_SM and the Level 2 algorithms. Most importantly, data assimilation requires unbiased estimates of modeled brightness temperatures with respect to the observations. When compared with SMOS observations, shown in Figure 10a, L-band modeled brightness temperatures typically exhibit a cold bias of several tens of Kelvin, shown in Figure 10c, if the parameter values for roughness and vegetation listed in the L2_SM_P ATBD are used (section 8). There are three likely reasons that contribute to this bias: (i) the Catchment model soil moisture climatology may be biased wet, (ii) the roughness and vegetation parameters from the L2_SM_P lookup table may not be appropriate for the global scale, and (iii) undetected and unmitigated low-level radio-frequency interference (RFI) may affect the SMOS brightness temperature observations. The impact of potential low-level RFI in SMOS observations cannot be established at this time. We will continue to monitor SMOS data product revisions and research results closely.

In a recent study, De Lannoy et al. (2012) calibrated select microwave radiative transfer parameters of the GEOS-5 modeling system such that the climatology of the modeled brightness temperatures matches that of the SMOS observations. The calibration was performed separately for each location, at the 36 km resolution in this study, and will be refined to 9 km for the L4_SM algorithm. For the calibration, we chose *not* to adjust the Catchment model soil moisture climatology. As discussed in section 2.2, there is not yet a consensus in the modeling or remote sensing communities about a global surface soil moisture climatology. Moreover, the Catchment model is an integral part of the NASA GMAO GEOS-5 system and a change in the soil moisture climatology through a model (or model parameter) modification would imply an undesirable divergence of the L4_SM modeling system from the GEOS-5 (or successor) system. The control vector for the calibration thus consists only of microwave radiative transfer model parameters related to microwave soil roughness, vegetation opacity, and vegetation scattering albedo; that is, the control vector does not include soil hydraulic or other land surface model parameters. The objective function penalizes deviations of the model climatology from the SMOS observations in terms of the long-term mean and the long-term standard deviation of brightness temperature for ascending, descending, H- and V-polarization, and multiple incidence angles. The objective function also penalizes deviations of the calibrated parameters from prior estimates.

After calibration, the climatological bias in brightness temperature is drastically reduced and less than 5 K for most of the globe (Figure 10e). The original bias in the time series variability (Figure 10d) is also reduced after calibration (Figure 10f). Unlike in the lookup table approach of the L2_SM_P algorithm, the calibrated microwave radiative transfer model parameters are not uniform within each vegetation class but still exhibit realistic spatial patterns (not shown). The modeled vegetation opacity after calibration also corresponds well to SMOS retrievals (not shown).

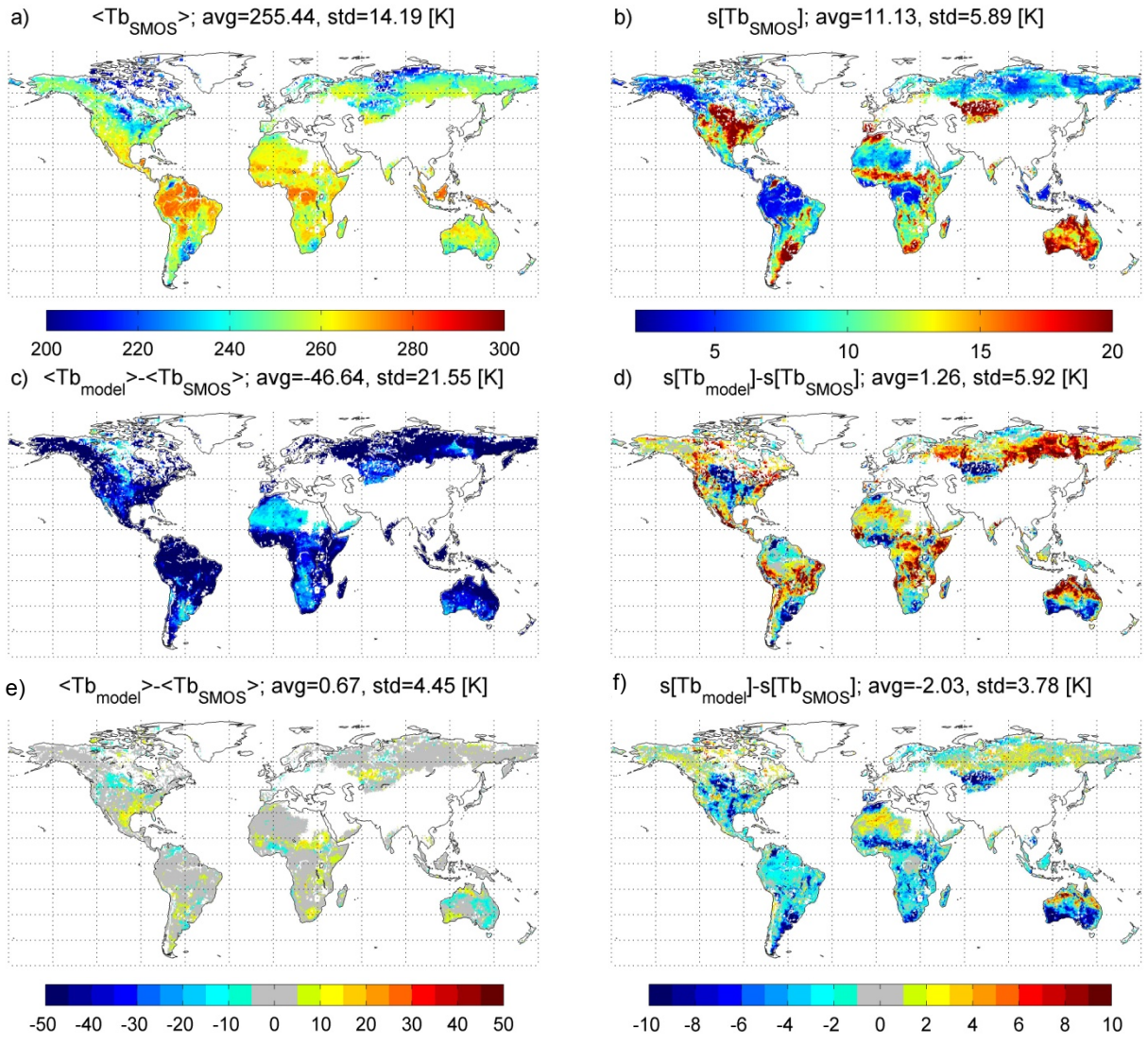

Figure 10. Time series (a) mean and (b) standard deviation of SMOS H-polarized Tb at 42.5° incidence angle during the validation period (1 Jan 2011 – 1 Jan 2012). (c) Mean difference between simulated Tb using microwave roughness and vegetation parameters from the L2_SM_P ATBD and SMOS observations. (d) Same as (c) but for difference of time series standard deviations. (e, f) Same as (c, d), respectively, but after calibration of microwave parameters. Titles indicate global averages (avg) and spatial standard deviations (std).

Again, the objective of the calibration is to minimize climatological bias. After calibration, the RMSE between SMOS estimates and the model remains as high as 10 K globally (not shown). Some of the remaining RMSE is related to random errors in the modeling system which will be addressed in the L4_SM algorithm through the assimilation of SMAP brightness temperature observations. Another contribution to the RMSE after calibration, however, is from seasonally varying biases. Figure 11 shows the temporal evolution in the differences between calibrated Tb simulations and SMOS observations as a function of latitude (averaged over longitude). The no-data periods at northern latitudes correspond to frozen conditions. There are clear seasonal patterns in the bias, most visibly at tropical and subtropical latitudes. Similar residual climatological differences are expected between modeled and SMAP-observed brightness temperatures and will be addressed prior to assimilation via the Z-score transformations of equation (5).

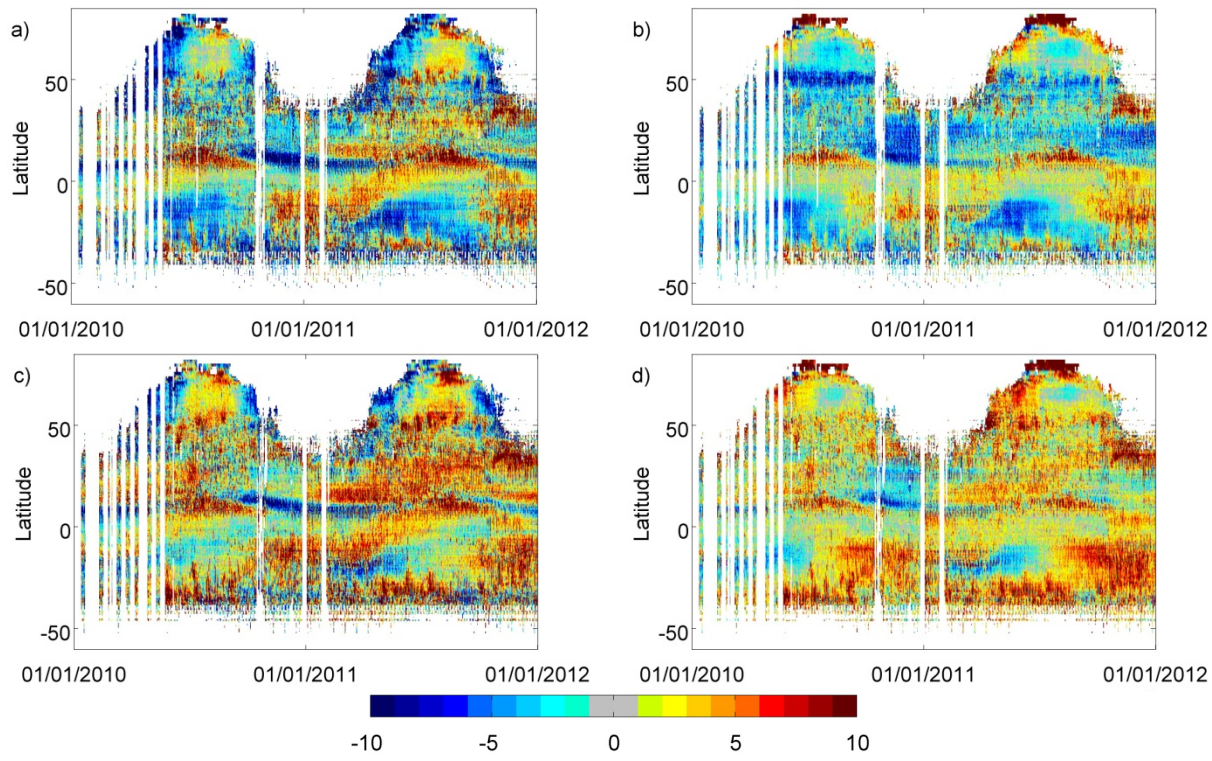

Figure 11. Hovmöller plots of (calibrated) model minus SMOS Tb [K] for 1 Jan 2011 – 1 Jan 2012, averaged over 6 incidence angles (32.5°, 37.5°, 42.5°, 47.5°, 52.5°, and 57.5°) for (a) ascending H-polarization, (b) ascending V-polarization, (c) descending H-polarization, and (d) descending V-polarization.

4.1.3c Surface meteorological data

The Catchment model is forced with surface meteorological data (including precipitation, downward shortwave radiation, downward longwave radiation, wind speed, near surface air temperature, near surface specific humidity, and air pressure). Select forcing inputs will be provided as part of the L4_SM research output (Table 6a; after appropriate interpolation in time and space, see below). The input forcing data stream will be provided by output from the global atmospheric analysis system at the NASA GMAO [*Rienecker et al*., 2008] and is based on the assimilation of a very large number (greater than 10^7 per day) of conventional and satellite-based observations of the atmosphere into a global atmospheric model. At the time of this writing, the resolution of the GMAO system is 0.25º by 0.3125º in latitude and longitude, respectively. By the time SMAP launches, the spatial resolution of these outputs is expected to be around 0.125º. The GMAO forcing data stream will be disaggregated to the SMAP 9 km EASE model grid with existing software. Furthermore, the forcing data are available as hourly averages or snapshots (depending on the variable) and will be interpolated to the land model time step with existing software.

Additional important corrections will be applied using gauge- and satellite-based estimates of precipitation. The corrections follow the procedure used by the Global Land Data Assimilation Systems (GLDAS) project [*Rodell et al*., 2003; http://ldas.gsfc.nasa.gov]. For example, the 2.5 degree pentad (5-day) dataset from the NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP; http://www.cdc.noaa.gov/cdc/data.cmap.html) is currently used for the initialization of NASA GMAO seasonal forecasts. For the SMAP L4_SM algorithm, the observations-based precipitation estimates will be downscaled to the hourly, 9 km scale of the model forcing using the disaggregation method described in [*Liu et al*., 2011; *Reichle et al*., 2011]. See Figure 2 and [*Liu et al*., 2011; *Reichle et al*., 2011] for further details regarding the impact of observations-based precipitation corrections.

The specific data source for the observations-based precipitation estimates will be determined closer to the launch of SMAP based on availability. At the time of this writing, global daily gauge-based estimates are provided by the NOAA Climate Prediction Center at a horizontal resolution of 0.5° with a latency of about 2 days (ftp://ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/). The Global Precipitation Measurement (GPM) mission is currently planning to produce only monthly average gauge-based estimates that would not satisfy the L4_SM latency constraints. In the extreme event that no global gauge-based precipitation product that satisfies L4_SM latency constraints should be available after the launch of SMAP, the L4_SM product could be generated using the precipitation forcing from the NASA GMAO data stream as is.

4.1.3d Land model initialization

The most appropriate way to initialize the Catchment model prognostic variables (Table 3) at the start of the assimilation period is to force the model (in ensemble mode) with meteorological data for a long time period prior to the assimilation, using forcing data extracted from the same sources used during the assimilation period [Rodell et al., 2005]. The Catchment land surface model will thus be forced with observation-based data for at least two years prior to the start of

the SMAP assimilation period. Memory of any poor initialization at the start of this "spin-up" period will be lost by the time the assimilation starts.

4.1.3e Data assimilation parameters

The key feature of the EnKF is that error estimates of the model-generated results are dynamically derived from an ensemble of model integrations. Each member of the ensemble experiences slightly perturbed instances of the observed forcing fields (representing errors in the forcing data) and is also subject to randomly generated noise that is directly added to the model prognostic variables (representing errors in model physics and parameters).

Time series of cross-correlated perturbation fields are generated and applied to selected meteorological forcing inputs and Catchment model prognostic variables. Collectively, these perturbations allow us to maintain an ensemble of land surface conditions that represents the uncertainty in the soil moisture states. An overview of the perturbation parameters is given in Table 4. Depending on the variable, normally distributed additive perturbations or lognormally distributed multiplicative perturbations are applied. The ensemble mean for all perturbations is constrained to zero for additive perturbations and to one for multiplicative perturbations. Moreover, time series correlations are imposed via a first-order auto-regressive model (AR(1)) for all fields. The perturbation fields are also spatially correlated (reflecting the 3d update step; section 4.1.2).

Perturbation	Additive (A) or Multiplicative (M)	Standard deviation	AR(1) time series correlation scale	Spatial correlation scale	Cross-correlation with perturbations in			
					P	SW	LW	TAIR
Precipitation (P)	M	0.5	24 h	50 km	n/a	-0.8	0.5	0
Downward shortwave (SW)	M	0.3	24 h	50 km	-0.8	n/a	-0.5	0.4
Downward longwave (LW)	A	20 W m^{-2}	24 h	50 km	0.5	-0.5	n/a	0.4
Air temperature (TAIR)	A	1 K	24 h	50 km	0	0.4	0.4	n/a
Catchment deficit	A	0.05 kg m^{-2}	3 h	25 km	n/a			
Surface excess	A	0.02 kg m^{-2}	3 h	25 km				
Surface temperature	A	0.2 K	12 h	25 km				

Table 4. Parameters for perturbations to meteorological forcing inputs and Catchment model prognostic variables. Perturbations are applied at hourly time steps.

For soil moisture, soil temperature, and brightness temperature, the dominant forcing inputs are precipitation, radiation, and air temperature, and we limit perturbations to these forcing

fields. Imperfect model parameters and imperfect physical parameterizations contribute to model errors. Such errors are represented through direct perturbations to model prognostic variables (that is, model parameter values such as porosity, soil hydraulic conductivity, vegetation opacity, single scattering albedo, etc. are not separately perturbed). The key prognostic variables of the Catchment model related to soil moisture and surface soil temperature are the surface excess, the root zone excess, the catchment deficit, and the surface temperature variables. Due to nonlinearities in the Catchment model, perturbations in the root zone excess typically lead to biases between the ensemble mean and the unperturbed control integration. We therefore limit the perturbations to the surface excess and the catchment deficit.

Cross-correlations are imposed on perturbations of the precipitation, radiation, and air temperature fields. At hourly and daily time scales, the meteorological forcing fields are ultimately based on output from atmospheric modeling and analysis systems and not on direct observations of surface precipitation and radiation. The cross-correlations are therefore motivated by the assumption that the atmospheric forcing fields represent a realistic balance between radiation, clouds, and precipitation. Under that assumption, a positive perturbation to the downward shortwave radiation tends to be associated with negative perturbations to the longwave radiation and the precipitation, and vice versa. The numbers for the imposed cross-correlation coefficients are motivated by an analysis of the cross-correlations between precipitation and radiation in the baseline forcing data sets from the Global Soil Wetness Project 2 [*International GEWEX Project Office*, 2002], and by the assumption that errors behave like the fields themselves.

Model and forcing errors are difficult to quantify at the global scale. The parameter values listed in Table 4 are largely based on experience. They are supported by earlier studies where model and forcing error parameters were calibrated in twin experiments [*Reichle et al.*, 2002b; *Reichle and Koster*, 2003] and by successful assimilation of SMMR, AMSR-E, ASCAT and ISCCP satellite observations [*Reichle et al.*, 2007; 2009a; *Liu et al.*, 2011; *Draper et al.*, 2012], suggesting that these values are acceptable. Additional calibration of the filter parameters with the optional adaptive filtering algorithm may further improve the assimilation results. A recent study by *Maggioni et al.* [2011] assessed the impact of rainfall error structure on soil moisture simulations by contrasting a complex satellite rainfall error model (SREM2D) to the standard rainfall error model of the GEOS-5 land assimilation system. Results show that perturbing satellite rainfall fields with the more complex SREM2D error model leads to improved spatial variability in the simulated soil moisture ensembles. However, when SREM2D was used in soil moisture assimilation, the improvement with respect to the standard GEOS-5 system were small [*Maggioni et al.,* 2012a, 2012b].

Observation error parameters for the assimilated SMAP brightness temperature Z-scores will be based on error estimates provided by the corresponding SMAP products. We anticipate that the ("instrument") measurement error standard deviation of SMAP brightness temperatures is ~1.3 K at 36 km (Table 1) and ~3.6 K at 9 km (L2_SM_AP ATBD, section 8), which will be converted to an equivalent dimension-less error in the assimilated Z-scores based on equations (5a) and (5b). Furthermore, the observation error standard deviation will include the contribution of "representativeness error" that is not yet included in the above-mentioned "instrument" error standard deviation. This representativeness error accounts, for example, for

the uncertainty associated with brightness temperature corrections for water bodies. The total observation error standard deviation (including the "instrument" error and the error of representativeness) will be determined during algorithm calibration (sections 4.1.4 and 4.2.4). We assume that observation errors are uncorrelated in time and space. The viability of this assumption will be checked against an analysis of the innovations and more complex error models will be developed if needed.

Again, the success of the assimilation system depends on the accurate specification of the model and observation error parameters. The improvements from data assimilation documented in Figures 2 and 3 suggest that the perturbation parameters listed in Table 4 are adequate, although not necessarily optimal. The adaptive filtering upscope option (section 4.1.2) may further enhance the assimilation performance. It may turn out during L4_SM algorithm development and calibration that the perturbation parameters should depend on space or on time. In this case, spatially distributed (or time-varying) fields of perturbation parameters will be included in the model parameters output file (Table 7a). Otherwise, the values of the perturbations parameters will be provided in a form similar to Table 4 as part of the documentation of the L4_SM data product.

4.1.4 Variance and Uncertainty Estimates

This section provides an overview of the **error budget** of the SMAP L4_SM product. Many sources of error contribute to the uncertainty in the L4_SM product. The key input uncertainties to the L4_SM algorithm are:

(i) Errors in the land model (a.k.a. "model error"), including errors in
   a. surface meteorological forcing data
   b. land model structure
   c. parameter values used in the land model
(ii) Errors in the assimilated SMAP Level 1-3 products (a.k.a. "observation error")
(iii) Errors in the specification of "model" and "observation" input error parameters

For the purposes of this discussion, the "land model" includes the Catchment model and the microwave radiative transfer model. As discussed above, uncertainty estimates are produced dynamically by the L4_SM algorithm along with estimates of geophysical parameters.

We expect that errors in the instantaneous surface soil moisture estimates from the L4_SM analysis product (SFMC_ANA, Table 6b) are roughly comparable to those of the L2_SM_AP product, and probably somewhat smaller, based on the additional information from the land surface model and its ancillary data (provided the data assimilation system uses properly calibrated input error parameters, sections 4.1.2 and 4.1.3). Given our lack of knowledge of the quality of future input data going into the assimilation system, however, the error structure of all the individual components of the L4_SM product cannot be specified a priori – particularly for subsurface soil moisture.

A "Level 4" Observing System Simulation Experiment (OSSE) framework has been developed to estimate the skill of outputs from the data assimilation system as a function of the

errors in the satellite retrievals and in the land model [*Reichle et al*., 2008a]. Note that by including a data assimilation system, the "Level 4" OSSE setup differs fundamentally from the "retrieval" OSSE setup [*Crow et al*., 2001, 2005a] that is used for SMAP Level 2 soil moisture products. The OSSE described here is based on the assimilation of soil moisture retrievals (as opposed to brightness temperatures), but we can still use it to get a rough estimate of the L4_SM error budget. Figure 12 shows a typical OSSE setup. The OSSE consists of a suite of synthetic data assimilation experiments that are based on integrations of two distinct land models, one representing "truth", and the other representing our flawed ability to model the true processes. The skill of the retrievals, model estimates, and assimilation products is measured in terms of the correlation coefficient R between the time series of the various estimates (expressed as anomalies relative to their long-term seasonal climatologies) and the assumed truth.

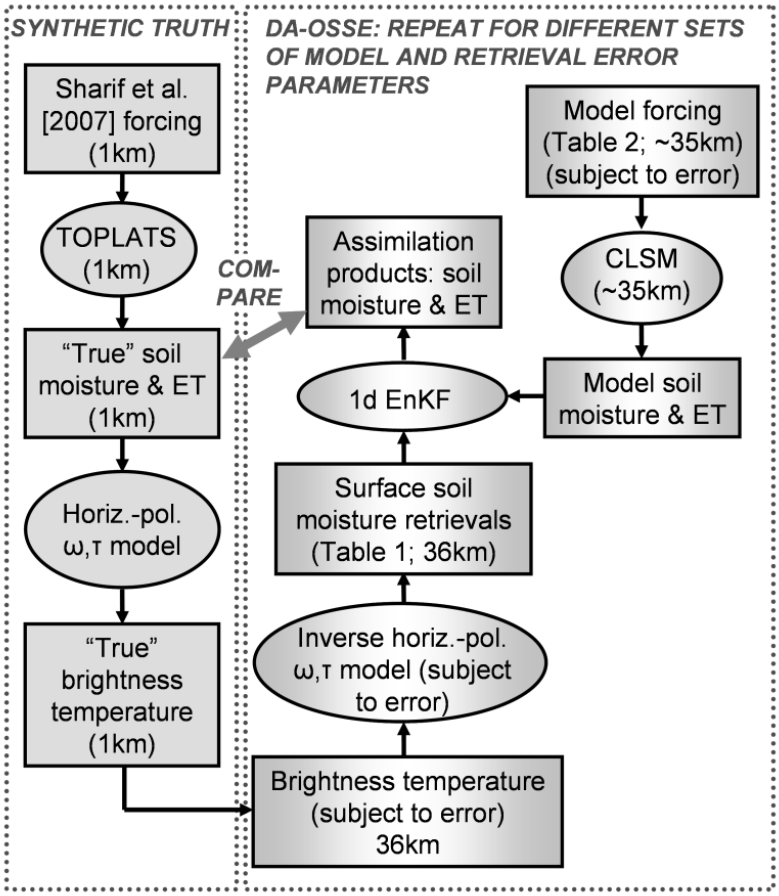

Figure 12. Flow diagram of a Level 4 Soil Moisture OSSE. From [*Reichle et al.*, 2008a].

Each assimilation experiment involves a unique combination of a retrieval dataset (with a certain level of skill, measured in terms of R) and a model scenario (with its own level of skill). We can thus plot two-dimensional surfaces of skill in the data assimilation products as a function of retrieval and model skill. For example, Figure 13a shows the two dimensional surface corresponding to the surface soil moisture product. As expected, the skill of the assimilation product generally increases with the skill of the model and the skill of the retrievals, for both surface (Figure 13a) and root zone (Figure 13b) soil moisture estimates. Except for very low model skill, the contour lines are more closely aligned with lines of constant model skill; that is, the skill of the assimilation product is more sensitive to model skill than to retrieval skill.

Figure 13 also shows skill improvement through data assimilation, defined as the skill of the assimilation product minus the skill of the model estimates (without assimilation). Specifically, Figures 13c and 13d show, for a given level of accuracy in the stand-alone model product, how much information can be added to the soil moisture products through the assimilation of satellite retrievals of surface soil moisture with a given uncertainty. Note that the skill of the surface and root zone soil moisture assimilation products always exceeds that of the model. As expected, the improvements in R through assimilation increase with increasing retrieval skill and decrease with increasing model skill. Perhaps most importantly, though, is that even retrievals of low quality contribute some information to the assimilation product, particularly if model skill is modest.

We can compare previously published skill levels with the results of Figure 13. For 23 locations across the contiguous United States with in situ observations appropriate for validation, *Reichle et al*. [2007; their Table 2] report, for surface soil moisture, average R values of 0.38, 0.43, and 0.50 for AMSR-E retrievals, Catchment model estimates, and their assimilation product, respectively. From the contours of Figure 13a we expect that for retrievals with R=0.38 and a model with R=0.43, the assimilation product would have skill of about R=0.50, which is indeed consistent with the AMSR-E result (indicated with a triangle in Figure 13a). For root zone soil moisture, *Reichle et al*. [2007] show that the assimilation of AMSR-E surface soil moisture retrievals also yields improvements, though these improvements fall somewhat short of those suggested by Figure 13b. Possible explanations include (i) the imperfect translation of information from the surface layer to the root zone in the data assimilation system and (ii) the fact that the in situ data used for validation of the AMSR-E result are themselves far from perfect (unlike the perfectly known truth of the synthetic experiment presented here). Figure 13 also includes the *Reichle et al*. [2007] results for assimilating retrievals from the historic Scanning Multichannel Microwave Radiometer (SMMR), which are similarly consistent with the contours. Note that R values for SMMR results are based on monthly mean data, and that the validating in situ data for the AMSR-E and SMMR results are not within the geographical domain of our synthetic experiment.

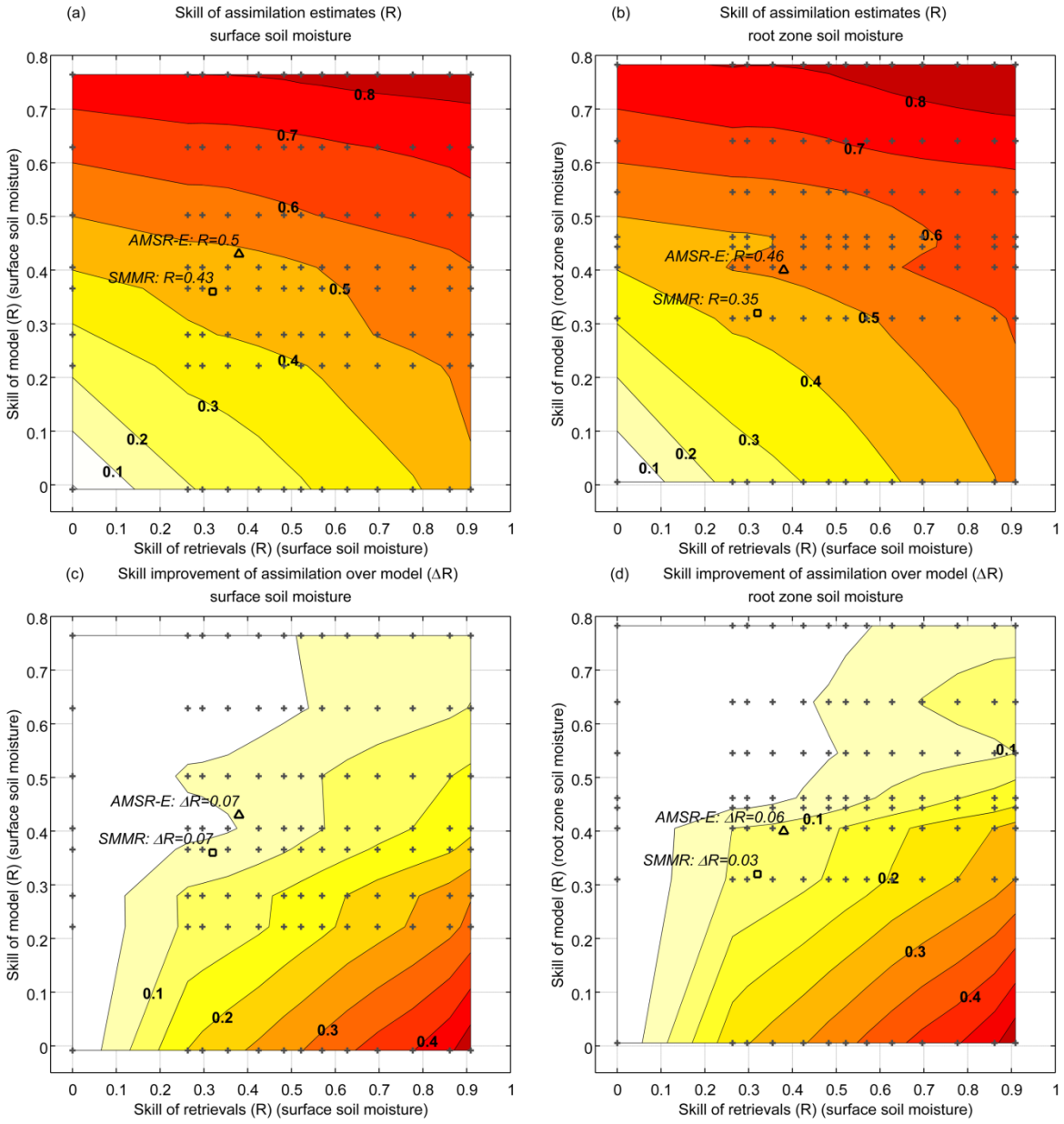

Figure 13. (a,b) Skill (R) and (c,d) skill improvement (ΔR) of assimilation product for (a,c) surface and (b,d) root zone soil moisture as a function of the (ordinate) model and (abscissa) retrieval skill. Skill improvement is defined as skill of assimilation product minus skill of model estimates. Each plus sign indicates the result of one 19-year assimilation integration over the Red-Arkansas domain. Also shown are results from *Reichle et al.* [2007] for (triangle) AMSR-E and (square) SMMR. From [*Reichle et al.*, 2008a].

Figure 14 shows the skill improvement (relative to the raw model product) for monthly mean evapotranspiration (ET) estimates from the data assimilation system. As expected, the assimilation of surface soil moisture retrievals contributes the most when retrievals are skillful and the model skill is poor.

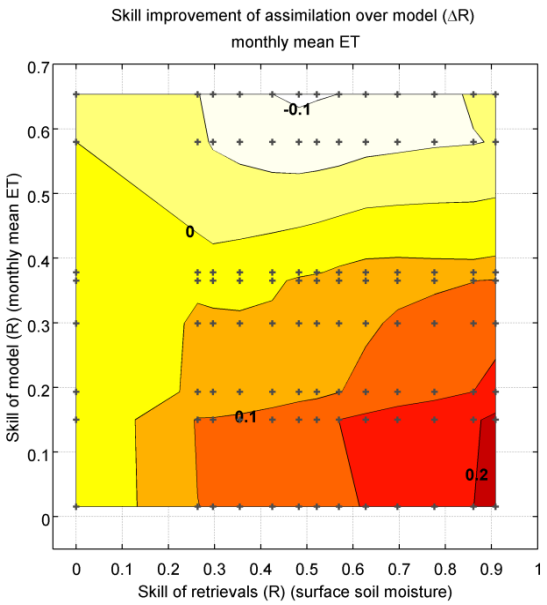

Figure 14. Skill improvement for monthly mean ET assimilation product. Abscissa, ordinate, and plus signs as in Figure 13. From [*Reichle et al.*, 2008a].

The results of Figure 13 which are based on synthetic observations have recently been corroborated by [*Draper et al.*, 2012] using ASCAT and AMSR-E observations. Figure 15 shows the skill increase (ΔR) relative to the open loop model from the single-sensor assimilation of ASCAT or AMSR-E surface soil moisture retrievals, as a function of the R value of the open loop model and of the assimilated (ASCAT or AMSR-E) retrievals. Since the R and ΔR values for the single-sensor assimilation of ASCAT or AMSR-E are generally similar, the results from the two experiments are combined. For a given combination of open-loop and observation skill, the skill gained through assimilation is slightly higher for root zone soil moisture (Figure 15b) than for surface soil moisture (Figure 15a). For both soil layers, assimilating observations with R no more than 0.2 below the open loop R (below the dashed lines in Figure 15) generally increased the soil moisture skill (i.e., $\Delta R>0$), with improvements up to $\Delta R\approx 0.4$ as the R value for the retrievals increases relative to that of the open loop.

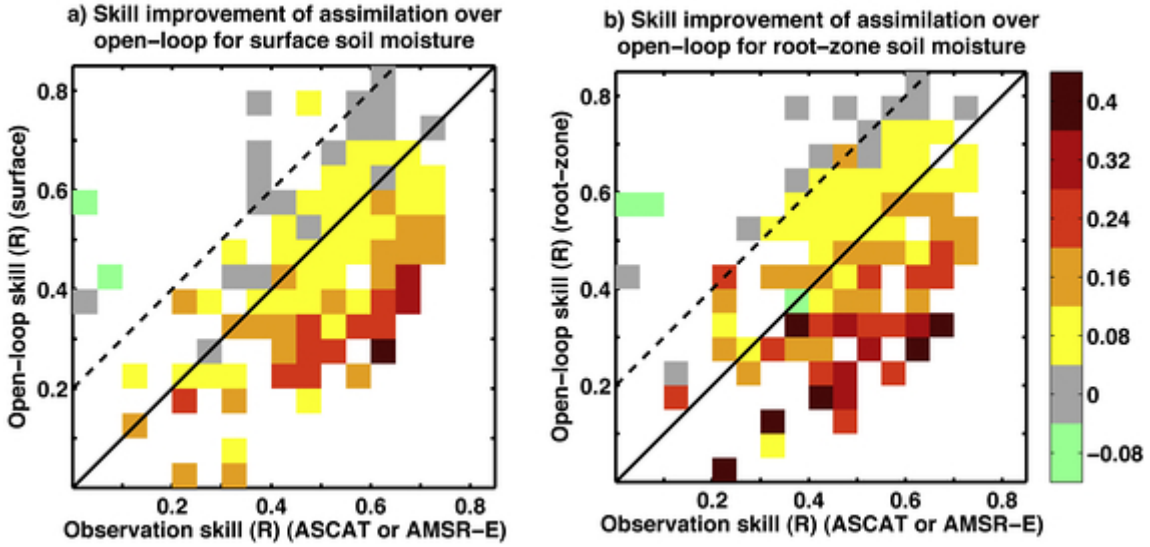

Figure 15. Skill improvement (ΔR) from assimilating either ASCAT or AMSR-E for (a) surface and (b) root zone soil moisture, as a function of the model open loop (no assimilation) and observation skill. Skill improvement is defined as the skill of the assimilation product minus the open loop skill, with skill based only on days with data available from both satellites. From [*Draper et al.,* 2012].

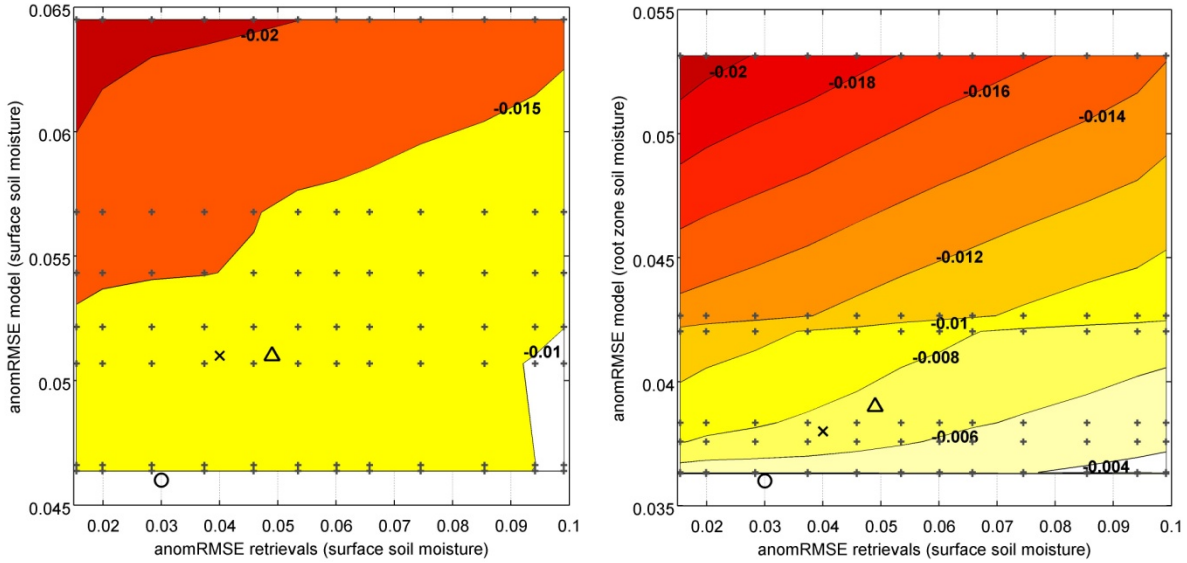

Figure 16. Skill improvement for (left) surface and (right) root zone soil moisture as a function of the (ordinate) model and (abscissa) retrieval skill based on data from [*Reichle et al.*, 2008a]. Skill is measured in terms of the anomaly RMSE ($m^3 m^{-3}$), and skill improvement is defined as skill of assimilation product minus skill of model estimates. Each plus sign indicates the result of one 19-year assimilation integration over the Red-Arkansas domain. Symbols indicate results for (triangle) AMSR-E, (x) SMAP L4_SM low skill scenario, and (o) SMAP L4_SM high skill scenario.

Based on the OSSE results by *Reichle et al*. [2008a], it is now straightforward to assess the expected uncertainty in the L4_SM soil moisture estimates. Figure 16 shows the improvement from the assimilation of surface soil moisture retrievals in terms of anomaly RMSE[4]. Negative numbers imply that the assimilation product is superior to the estimates from the land model alone. Triangles indicate the skill improvement from assimilating AMSR-E retrievals (based on results updated from [*Reichle et al*., 2007]). For SMAP, we now assume two separate scenarios, a low skill scenario with higher retrieval and model error levels, and a high skill scenario with correspondingly lower error levels (Table 5). Based on these assumed skill scenarios, we can use Figures 14 and 16 to read off the expected error levels and skill improvements of the L4_SM surface and root zone soil moisture estimates. The resulting estimates are tabulated in the last two columns of Table 5. Based on the OSSE, we roughly expect that the assimilation of SMAP observations improves the anomaly RMSE of L4_SM surface soil moisture estimates by ~0.01 $m^3 m^{-3}$ over the corresponding Catchment model estimates. For L4_SM root zone soil moisture estimates we expect anomaly RMSE improvements of ~0.005 $m^3 m^{-3}$. It is important to note, however, one key limitation of this error analysis. The model skill estimates were determined in a region, the contiguous United States, where in situ observations were available. Along with in situ soil moisture observations, the US also offers good coverage of precipitation gauges (Figure 17), which tends to ensure good model skill. Improvements from soil moisture assimilation

[4]Anomaly RMSE is the RMSE computed after removing the long-term seasonal climatology from the time series of the validation data and the model (or assimilation) estimates.

should be larger for regions with less reliable precipitation data, including most of South America, Africa, Asia, and Australia (Figure 17).

	Skill scenario	Model[1,2]	L4_SM[2]	|Δ|
Expected anomaly RMSE [m^3/m^3]				
Surface soil moisture	High	0.046	0.035	0.012
	Low	*0.051*	*0.038*	*0.012*
Root zone soil moisture	High	0.036	0.031	0.005
	Low	*0.038*	*0.031*	*0.007*
Expected anomaly R				
Surface soil moisture	High	0.63	0.71	0.08
	Low	*0.41*	*0.54*	*0.13*
Root zone soil moisture	High	0.55	0.63	0.08
	Low	*0.46*	*0.59*	*0.13*

[1]Source: USDA/SCAN results.
[2]Source: OSSE results.

Table 5. Expected anomaly RMSE and R for L4_SM soil moisture estimates based on the OSSE data from [*Reichle et al.*, 2008a]. Assumed retrieval skill levels are based on SMAP measurement requirements. Model skill estimates are based on validation of Catchment model integrations against USDA/SCAN in situ observations.

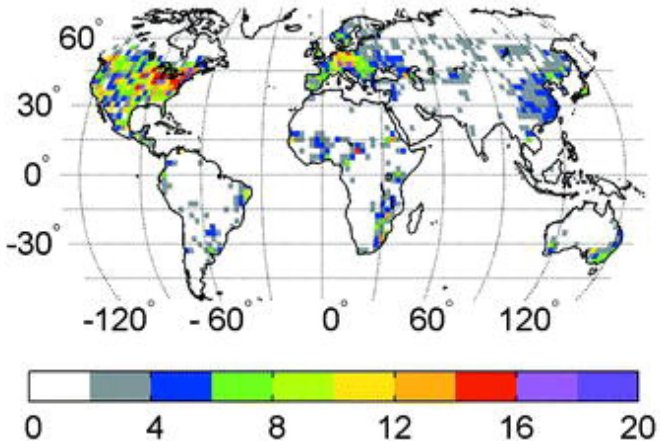

Figure 17. Average monthly number of rain gauges per 2.5° grid cell (1979-1987; from [*Reichle et al.,* 2004]).

In summary, the general OSSE framework permits detailed and comprehensive error budget analyses for data assimilation products. The framework can be used, for example, to study specific trade-offs in ancillary data requirements, assessing the impact of each on the quality of the end-product that will be used in science and applications. Conceptually, extending the OSSE to global scales or higher-resolution is straightforward, but computational costs may prohibit an analysis that is comparable to that described in Figures 14-16. Another straightforward extension of the OSSE framework is to assimilate microwave brightness temperatures directly (as opposed to surface soil moisture retrievals) and then examine how uncertainties in the

retrieval process may be mitigated through use of a priori information from the land surface model, notably surface soil temperature. Finally, note that the OSSE results, just like the accuracy of the L4_SM data product, depend on the realism of the underlying model and observation error structure and parameters (see section 4.1.3).

4.2 Practical Considerations

4.2.1 Numerical Computation Considerations

Computing and storage requirements for the SMAP L4_SM product are very manageable when compared to those of global atmospheric data assimilation systems but do exceed the resources of typical desktop or small cluster environments that are maintained by individual research groups.

The SMAP L4_SM product will be developed on the Linux-based Discover cluster at the NASA Center for Climate Simulation (http://www.nccs.nasa.gov) located at GSFC. The CPU requirements will scale approximately linearly with the number of ensemble members used in the ensemble Kalman filter and will also depend on the degree of horizontal correlation assumed in the 3d assimilation process (section 4.1.2). Implementation of the L4_SM algorithm targets a processing speed of at least 10X (that is, processing 10 data days per day) to facilitate reprocessing. Typical jobs will require 128-256 processors.

Minimum on-line (hard-drive) storage requirements would be on the order of a few hundred GB per data month. Total storage requirements for the archive are on the order of tens of TB. Detailed information can be found in SMAP data product specification documents that are maintained by the SMAP Project. These estimates are a first cut and cover a time period that is long with respect to the development of computing technology. As the L4_SM product definitions will be refined by the SMAP team and as computing technology improves, computing requirements may change accordingly.

The L4_SM algorithm will require inputs of SMAP Level 1-3 products that will be generated at JPL. After L4_SM product generation at GSFC, the L4_SM product will be transmitted for storage to the DAAC (National Snow and Ice Data Center, section 2.3.2). Bandwidth for data transfer will thus be an important consideration.

4.2.2 Programming/Procedural Considerations

The L4_SM algorithm will be written in Fortran 90 for use on a multi-processor Linux cluster (section 4.2.1).

4.2.3 Ancillary Data Availability/Continuity

NASA HQ and the GMAO are committed to continued production of the GEOS-5 (or successor) data stream that provides surface meteorological input data for the L4_SM algorithm.

4.2.4 Calibration and Validation

Validation of the L4_SM product is primarily against available in situ observations. Consistency of internal data assimilation diagnostics as well as consistency with related observations, such as precipitation measurements, provides supplemental validation. Validation to mission requirements will be consistent with the requirements of the *SMAP Science Data Calibration and Validation Plan* (section 8) and focus on the land areas specified in the *SMAP Level 1 Requirements and Mission Success Criteria* document (excluding regions of snow and ice, mountainous topography, open water, urban areas, and vegetation with a total water content greater than 5 kg m^{-2}). Outside of these areas, validation will be attempted to the extent possible.

4.2.4a Validation with in situ observations

Validation issues specific to the L4_SM product

The L4_SM product merges SMAP observations with ancillary data (including precipitation) through integration with a land surface model. In this process, land models perform complex, nonlinear energy and water balance calculations across large spatial scales. However, a global land model integration at grid scales consistent with the resolutions of the SMAP instruments cannot resolve the fine scale structure of spatial soil moisture variations (~tens of meters) that help determine spatially-averaged energy and water fluxes. Moreover, we have imperfect knowledge of the global distributions of land surface characteristics (for example, soil texture, depth-to-bedrock, and emissivities, which also vary at sub-SMAP scales) that control simulated subsurface flow as well as microwave emission and backscatter. Thus, by necessity, the soil moisture and brightness temperature estimates generated by land models reflect simplifying assumptions. Consequently, the mean values and variances of soil moisture estimates differ between land models and in-situ (point-scale) observations (Section 2.2).

These limitations, however, in no way negate the usefulness of combining SMAP observations with meteorological data through data assimilation. While a land model product derived from observed meteorological forcing may not match the brightness temperature observed by SMAP (or the Level 2 soil moisture retrieved from SMAP) in absolute magnitude, the *time variation* of the land model and satellite products must be consistent. Both the observed (or downscaled) and modeled brightness temperature, for example, should be anomalously high following an extended dry period. The optimal merging of SMAP and land model data focuses on this shared and fundamental information content. A similar argument applies to the soil moisture estimates from satellite retrievals, model/assimilation output, or in situ measurements. By knowing the temporal moments (or climatology) of the datasets, one dataset can be "scaled" (for example, via cdf matching) to be fully consistent with the climatology of the other, at which point the two datasets can be merged together with confidence or compared directly for purposes of validation.

For certain applications, such as the initialization of soil moisture reservoirs in atmospheric forecasting systems, the absolute error in the soil moisture is not necessarily relevant [*Crow et al.*, 2005b]. Since scaling of soil moisture data is almost always required prior to their use in model-based applications, time-invariant biases in the moments of the L4_SM product often become meaningless. For model applications, the temporal correlation of soil moisture estimates with independent observations may therefore be a more relevant validation metric [*Entekhabi et al.,* 2010a]. By focusing on time variations, evaluation problems stemming from the inconsistency between point and area-averaged quantities are, to some extent, ameliorated.

Validation criteria

The L4_SM *soil moisture* estimates will be *validated* against "verified" in situ soil moisture measurements from operational networks and dedicated SMAP field experiments (as summarized in the *SMAP Science Data Calibration and Validation Plan*, section 8). Specifically, the L4_SM surface and root zone soil moisture estimates will satisfy the following criterion:

> **RMSE ≤ 0.04 m^3 m^{-3}** within the data masks specified in the *SMAP Level 2 Science Requirements* (section 8; excluding regions of snow and ice, frozen ground, mountainous topography, open water, urban areas, and vegetation with water content greater than 5 kg m^{-2}),

where RMSE is computed after removing long-term mean bias. This criterion applies to the L4_SM instantaneous surface soil moisture outputs in the climatology of the L2_SM_AP retrievals (SFMC_ANA_L2CLIM; Table 6b) and the L4_SM instantaneous root zone soil moisture estimates (RZMC_ANA_PRCNTL; Table 6b) after conversion from percentile to volumetric units (consistent with the climatology of the Catchment model) based on the cdf parameters provided in the "clim" output file (Table 7b).

L4_SM output fields other than soil moisture are provided as *research products* (including surface meteorological forcing variables, surface soil temperature, evaporative fraction, net radiation, etc.; section 2.3.2). Specifically, L4_SM land surface temperature and flux estimates will be evaluated against in situ observations where possible (see below).

As part of the validation process, additional metrics (including bias, RMSE, anomaly RMSE, and R values) will be computed for the L4_SM output fields to the extent possible. This includes computation of the metrics outside of the limited geographic area for which the 0.04 m^3 m^{-3} validation criterion is applied. The additional metrics will be made available to the SMAP project for the SMAP Cal/Val report. For the computation of the *anomaly* metrics, the seasonal cycle of the raw data (L4_SM product and validating in situ observations) will be estimated for each location by computing, for each day of the year (DOY), a climatological value of soil moisture. It is important to obtain as many years of validating in situ measurements as possible *prior* to the launch of SMAP. We estimate that 5 years of measurements would be sufficient for the anomaly metrics, although this requirement could perhaps be relaxed through the ergodic substitution of variability in space for variability in time to estimate the climatology. Note that the Catchment model climatology can be computed prior to the launch of SMAP.

Available in situ observations

In the following, we give only a brief overview of the most relevant in situ observations that can be used for validation. The *SMAP Science Data Calibration and Validation Plan* (section 8) provides a detailed and comprehensive description of the available validation data sets, including a complete listing of "verified" in situ observations.

High-quality in situ observations of *surface* soil moisture that approximate the scale of satellite and model estimates through distributed sensor networks are available for select watersheds [*Jackson et al.*, 2010]. Four USDA Agricultural Research Service (ARS) CalVal watersheds with on-going data collection and records going back to 2002 are shown in Figure 18. Instrumentation of additional watersheds that cover a range of climate and vegetation conditions is planned for SMAP as discussed in the *SMAP Science Data Calibration and Validation Plan* (section 8).

The USDA Soil Climate Analysis Network (SCAN; http://www.wcc.nrcs.usda.gov) provides hourly, ground-based observations of soil moisture and soil temperature (Figure 18). Measurements are taken at depths of 5 cm, 10 cm, 20 cm, 51 cm, and 102 cm (wherever possible). Soil moisture is estimated with a device that measures the dielectric constant of the soil. The SCAN archive requires extensive and strict quality control based on close inspection of all data points by the user [*Reichle et al*., 2007; *Liu et al.*, 2011]. In many cases, there are spikes in the soil moisture time series resulting from individual soil moisture data values that are obviously too large (well above any realistic porosity) or too small (sometimes even negative). Soil moisture observations must be excluded whenever the corresponding soil temperature measurements indicate that the soil is frozen. There are also obvious inconsistencies in the time series that might have been caused by changes in sensor calibration (recorded in the accompanying documentation) or by apparent detachment of the sensor from the soil matrix. Recent revisions of the archived data and of the processing algorithm have addressed some of these concerns. Nevertheless, whenever there is any doubt about the validity of a data point or a part of the time series, it is recommended that the measurements in question be excluded and that no data be filled in or interpolated. For example, *Liu et al.* [2011] were able to use about 50 SCAN stations for validation (Figure 18) and *Draper et al.* [2012] used a total of 66 stations in the U.S. and Australia.

In situ soil moisture profile observations are also available from the Oklahoma Mesonet (http://www.mesonet.org/). By the time SMAP launches, additional soil moisture networks may provide validating observations. The NOAA National Climatic Data Center is currently installing a network of soil moisture sensors similar to the SCAN network across the US with the purpose of establishing a long-term climate data record. It has been demonstrated that signals from the Global Positioning System of satellites can be used to infer soil moisture [*Larson et al*., 2008]. Finally, a network that measures soil moisture via cosmic-ray neutron scattering signals is currently being implemented [*Zweck et al.*, 2008]. The above discussion only touches on a few key data sets. Additional data sets are available from the International Network for in Situ Soil Moisture Data (http://www.ipf.tuwien.ac.at/insitu) [*Dorigo et al.*, 2011]. Again, the *SMAP Science Data Calibration and Validation Plan* (section 8) provides a detailed and comprehensive description of the available soil moisture validation data sets.

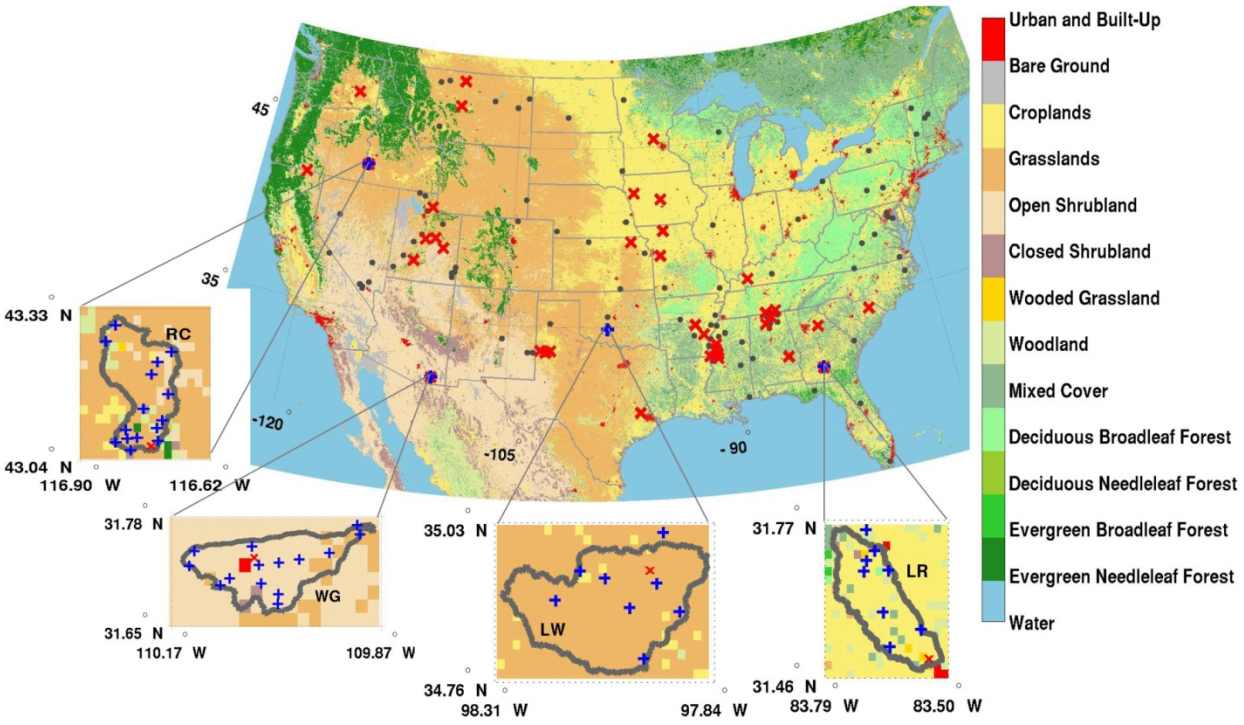

Figure 18. Location of (red crosses) SCAN sites and (blue plus signs) ARS CalVal watersheds. SCAN sites not used for validation in [*Liu et al.*, 2011] are marked with dots (see Figure 2 for results from [*Liu et al.*, 2011]). The background shows the MODIS land cover product based on UMD classification (from http://duckwater.bu.edu/lc/mod12q1.html) at ~2 km resolution. From [*Liu et al.*, 2011].

The availability of land surface flux data for validation is very limited. A comparably large collection of such data is provided free of charge by the Coordinated Energy and Water Cycle Observations Project (CEOP; http://www.ceop.net) and will be used to validate the data assimilation products. From 1 October 2002 through 31 December 2004, for example, 24 CEOP reference sites, located mostly in Kansas and Oklahoma, provide hourly surface flux data that are sufficient for validation [*Reichle et al.*, 2010]. A more comprehensive set of surface flux data is available through Fluxdata.org (http://www.fluxdata.org).

4.2.4b Validation with internal assimilation diagnostics

Relative to the coverage of the satellite and model soil moisture estimates, few in situ data are available. The soil moisture data assimilation system produces internal diagnostics that will be used to indirectly validate its output. Specifically, the statistics of appropriately normalized innovations will be examined [*Reichle et al.*, 2007; see also discussion of adaptive filtering in section 4.1.2].

4.2.4c Validation with high-quality, independent precipitation observations

Validation with in situ soil moisture observations is difficult because there are few long-term station observations and because there is a mismatch between the point-scale of the in situ measurements and the distributed (9 km) scale of the L4_SM product. Compared to ground-

based soil moisture probes, rain gauges are inexpensive, easy to maintain, and have already been widely installed over vast continental regions. Moreover, variability in daily rainfall accumulations occurs at spatial scales that are typically coarser than the fine-scale (potentially < 10 m) variability of soil moisture. Because errors in soil moisture are primarily a result of errors in precipitation, and because precipitation observations are more abundant and reflective of more appropriate scales, gauge-based precipitation observations can be used for an indirect evaluation of soil moisture estimates.

Crow and Zhan [2007] developed a data assimilation-based approach for evaluating surface soil moisture retrievals that effectively substitutes rain gauge measurements for ground-based soil moisture observations. The approach is based on evaluating the correlation coefficient between antecedent rainfall error and analysis increments that are produced by the soil moisture assimilation system. The use of rain gauge observations expands potential soil moisture validation locations from isolated sites (Figure 18) to continental-scale regions over which high-quality rain gauge measurements are available. A modified form of this approach was used to evaluate the added value of AMSR-E based soil moisture retrievals for root zone soil moisture monitoring within the continental United States [*Bolten et al.*, 2009] and extended to a quasi-global evaluation system by *Crow et al.* [2010]. The approach will be applied to evaluate the increments that are produced by the L4_SM algorithm.

4.2.5 Quality Control and Diagnostics

Quality control is an integral part of the soil moisture assimilation system. At least two kinds of quality control (QC) measures are needed. The first set of QC steps is based on the flags that are provided with the SMAP observations. We will assimilate only SMAP brightness temperature data that have favorable flags for soil moisture estimation (for example, acceptable vegetation density, no rain, no snow cover, no frozen ground, no RFI, sufficient distance from open water).

The second set of QC steps are additional "online" rules that exclude SMAP observations from assimilation in the EnKF (soil moisture) update whenever the land surface model indicates that (i) rain is falling, (ii) the soil is frozen, or (iii) the ground is fully or partly covered with snow. Note also that the assimilation system will typically provide some weight to the model background and thus buffers the impact of anomalous observations that may slip through the flagging process.

4.2.6 Exception Handling

See section 4.2.5.

4.2.7 Interface Assumptions

Inputs

The ancillary input data described in section 4.1.3 are unique to the L4_SM product and will be obtained from the GMAO, except for the precipitation observations that are used to correct the GMAO precipitation estimates. These observations will likely be obtained from the NOAA Climate Prediction Center (section 4.1.3). In addition to the ancillary data, the SMAP L1C_TB, L2_SM_AP, and L3_FT_A products are required for the baseline algorithm. The SMAP L4_SM algorithm requires that the input SMAP Level 1-3 products contain, at a minimum, swath or gridded observations with corresponding latitude, longitude, and time information. The SMAP Level 1-3 products will be provided on the SMAP Earth-fixed nested global grids and also contain the quality control flags discussed in section 4.2.5.

The L4_SM algorithm also requires estimates of the measurement errors in the L1C_TB and L2_SM_AP brightness temperatures. If such error estimates are provided with the Level 1 and Level 2 products, they will be used dynamically in the L4_SM algorithm. If not provided by the input products, default error estimates will be developed as part of the L4_SM calibration (see also section 4.1.3).

The input SMAP Level 1-3 products listed above must be available with latencies of at least two days less than the corresponding latency of the L4_SM product (specified in section 2.3.2) to allow sufficient time for L4_SM processing.

Outputs

The baseline SMAP Level 4 Carbon Net Ecosystem Exchange (L4_C) product relies on the soil moisture and soil temperature estimates from the L4_SM product. Once generated, the L4_SM product is transferred to the DAAC for permanent archival.

4.2.8 Test Procedures

Before the SMAP launch, the L4_SM algorithm will be tested globally, to the extent possible, with satellite observations from the precursor missions discussed in section 2.2. SMOS, a passive microwave sensor operating at L-band and launched successfully on 2 Nov 2009, will play a key role. In each case, the outcome of the tests will be assessed by validating the assimilation estimates against in situ observations from existing networks and by ensuring the consistency of internal diagnostics (section 4.2.4). Additional development and testing will be conducted in the context of OSSEs (section 4.1.4).

4.2.9 Algorithm Baseline Selection

The selection of the final algorithm will be part of the testing with precursor satellite observations described in section 4.2.8. If the validation process of a test scenario indicates that an option algorithm provides an improvement in the product that justifies the added complexity,

the option algorithm will be used. The L4_SM algorithm will be built from the existing GEOS-5 land assimilation system (sections 2.2 and 4.1.1). The following steps roughly describe the development path:

- Implement forward radiative transfer model. [*complete*]
- Calibrate climatology of forward radiative transfer model to SMOS observations. [*in progress*]
- Establish model skill based on observations from SCAN and USDA core validation sites. [*complete; will be updated regularly with the most recent data*]
- Test brightness temperature assimilation (using synthetic, AMSR-E, SMOS, and Aquarius observations). [*in progress*]
- Implement assimilation of freeze-thaw product (L3_FT_A). [*in progress*]
- Implement and test upscope options. [*pending availability of resources*]

5. Acknowledgments

The authors would like to thank the members of the SMAP Science Definition Team, the SMAP Algorithm Development Teams, Gianpaolo Balsamo, Matthias Drusch, Steven Margulis, Joaquin Munoz-Sabater, and Eric Wood for valuable discussions and feedback.

6. Prototype Data Product Specifications

Tables 6a and 6b provide an overview of the geophysical fields to be included in the L4_SM data product. The data product will consist of two sets of files. One set of files is for time-average fields, including soil moisture, soil temperature, snow water equivalent, land surface turbulent fluxes, and land surface forcing variables such as precipitation and radiation. The second set is for instantaneous ("snapshot") output variables and includes forecast and analysis soil moisture as well as forecast and observed brightness temperatures. Detailed data product specifications including headers and metadata are under development.

Tables 7a and 7b provide an overview of the time-invariant input variables. Table 7a lists the time-invariant input parameters for the Catchment land surface model and its radiative transfer module, including soil and vegetation parameters. Table 7b lists the climatological parameters for brightness temperature and soil moisture. The climatological parameters for brightness temperature are broken down by time-of-day and calendar month.

Data group name	**SMAP Level 4 Surface and Root Zone Soil Moisture time average output**	
File name	SMAP_L4_SM_tavg_vVVV_YYYYMMDD_HHMMz.h5 (if stored in separate file)	
File format	netcdf4/hdf5	
Characteristics	Time averaged at native resolution	
Spatial dimensions	Longitude: **3852**, latitude: **1632** (SMAP 9 km EASE grid)	
Times	**1:30z, 4:30z, 7:30z, 10:30z, 13:30z, 16:30z, 19:30z, 22:30z**	
Variable name	**Description**	**Units**
SFMC_L2CLIM	Surface soil moisture (0-5 cm; consistent with L2_SM_AP climatology)	$m^3 m^{-3}$
SFMC_PRCNTL	Surface soil moisture (0-5 cm; percentile units)	[-]
RZMC_PRCNTL	Root zone soil moisture (0-100 cm; percentile units)	[-]
PRMC_PRCNTL	Total profile soil moisture (0-CLSM_DZPR$^+$; percentile units)	[-]
TSURF	Surface temperature (canopy and top soil layer of depth CLSM_DZTS$^+$)	K
TSOIL1	Soil temperature in layer 1 of soil heat diffusion model$^+$	K
TSOIL2	Soil temperature in layer 2 of soil heat diffusion model$^+$	K
TSOIL3	Soil temperature in layer 3 of soil heat diffusion model$^+$	K
TSOIL4	Soil temperature in layer 4 of soil heat diffusion model$^+$	K
TSOIL5	Soil temperature in layer 5 of soil heat diffusion model$^+$	K
TSOIL6	Soil temperature in layer 6 of soil heat diffusion model$^+$	K
SNOMAS	Snow mass	$kg\ m^{-2}$
SNODP	Snow depth	m
EVLAND	Evapotranspiration from land (excluding open water and permanent ice)	$kg\ m^{-2}\ s^{-1}$
RUNOFF	Overland (surface) runoff	$kg\ m^{-2}\ s^{-1}$
BASEFLOW	Baseflow	$kg\ m^{-2}\ s^{-1}$
SNOMLT	Snowmelt	$kg\ m^{-2}\ s^{-1}$
FRSAT	Fractional saturated area	[-]
FRUNSAT	Fractional unsaturated but non-wilting area	[-]
FRWLT	Fractional wilting area	[-]
FRSNO	Fractional snow-covered area	[-]
SHLAND	Sensible heat flux from land (excluding open water and permanent ice)	$W\ m^{-2}$
LHLAND	Latent heat flux from land	$W\ m^{-2}$
GHLAND	Downward ground heat flux at base of top soil layer	$W\ m^{-2}$
SWLAND	Net downward shortwave flux over land	$W\ m^{-2}$
PRECTOT*	Total surface precipitation	$kg\ m^{-2}\ s^{-1}$
PRECCONV*	Convective surface precipitation	$kg\ m^{-2}\ s^{-1}$
PRECSNO*	Surface snow fall	$kg\ m^{-2}\ s^{-1}$
LWGAB*	Absorbed (downward) longwave radiation at the surface	$W\ m^{-2}$
SWGDN*	Downward shortwave flux incident on the surface	$W\ m^{-2}$
PS*	Surface pressure	Pa
HLML*	Center **h**eight of **l**owest atmospheric **m**odel **l**ayer	m
TLML*	Air temperature at HLML	K
QLML*	Air specific humidity at HLML	$kg\ kg^{-1}$
ULML*	Eastward wind at HLML	$m\ s^{-1}$
VLML*	Northward wind at HLML	$m\ s^{-1}$
LAI*	Leaf area index	[-]
GRN*	Greenness	[-]

$^+$*See Table 7a for corresponding layer depths.*

Table 6a. L4_SM time average output. For example, time averages with 1:30z time stamp are averages from 0z to 3z. Outputs are ensemble mean values. *Ancillary inputs to the L4_SM algorithm (provided along with L4_SM outputs after some processing; section 4.1.3).

Data group name	**SMAP Level 4 Surface and Root Zone Soil Moisture analysis update output**	
File name (if stored in separate file)	SMAP_L4_SM_aupd_vVVV_YYYYMMDD_HHMMz.h5	
File format	netcdf4/hdf5	
Characteristics	Instantaneous at native resolution	
Spatial dimensions	Longitude: **3852**, latitude: **1632** (SMAP 9 km EASE grid)	
Times	**0z, 3z, 6z, 9z, 12z, 15z, 18z, 21z**	
Variable name	**Description**	**Units**
TIME	Approximate time of SMAP overpass	TBD
TBHCOMP_RESFLAG‡	Flag indicating effective resolution of H-pol brightness temperature composite fields (TBHCOMP_OBS, TBHCOMP_FCST, etc.): 1=36 km, 2=9 km	[-]
TBVCOMP_RESFLAG‡	Flag indicating effective resolution of V-pol brightness temperature composite fields (TBVCOMP_OBS, TBVCOMP_FCST, etc.): 1=36 km, 2=9 km	[-]
TBHCOMP_OBS‡	Composite resolution observed (L2_SM_AP or L1C_TB) H-pol brightness temperature+ &	K
TBVCOMP_OBS‡	Composite resolution observed (L2_SM_AP or L1C_TB) V-pol brightness temperature+ &	K
TBHCOMP_OBS_ZSCORE‡	Z-score of TBHCOMP_OBS+ (assimilated)	[-]
TBVCOMP_OBS_ZSCORE‡	Z-score of TBVCOMP_OBS+ (assimilated)	[-]
TBHCOMP_OBS_ZSCORE_ERRSTD‡	Observation error std-dev for TBHCOMP_OBS_ZSCORE+	[-]
TBVCOMP_OBS_ZSCORE_ERRSTD‡	Observation error std-dev for TBVCOMP_OBS_ZSCORE+	[-]
FT_L3_FT_A‡	Observed (L3_FT_A) freeze-thaw state& (assimilated)	[-]
TBHCOMP_FCST‡	Composite resolution Catchment model forecast 1.41 GHz H-pol brightness temperature#	K
TBVCOMP_FCST‡	Composite resolution Catchment model forecast 1.41 GHz V-pol brightness temperature#	K
TBHCOMP_FCST_ZSCORE‡	Z-score of TBHCOMP_FCST#	[-]
TBVCOMP_FCST_ZSCORE‡	Z-score of TBVCOMP_FCST#	[-]
TBHCOMP_FCST_ZSCORE_ENSSTD‡	Ensemble std-dev for TBHCOMP_FCST_ZSCORE#	[-]
TBVCOMP_FCST_ZSCORE_ENSSTD‡	Ensemble std-dev for TBVCOMP_FCST_ZSCORE#	[-]
SFMC_FCST_PRCNTL	Catchment model forecast surface soil moisture (0-5 cm)	[-]
RZMC_FCST_PRCNTL	Catchment model forecast root zone soil moisture (0-100 cm)	[-]
PRMC_FCST_PRCNTL	Catchment model forecast total profile soil moisture (0-CLSM_DZPR)	[-]
TSURF_FCST	Catchment model forecast surface temperature	K
TSOIL1_FCST	Catchment model forecast first-layer soil temperature	K

*[**Table 6b**, continues on next page]*

[continued from previous page]

SFMC_ANA_L2CLIM	Analysis surface soil moisture (0-5 cm; consistent with L2_SM_AP climatology)	$m^3 m^{-3}$
SFMC_ANA_PRCNTL	Analysis surface soil moisture (0-5 cm; percentile units)	[-]
RZMC_ANA_PRCNTL	Analysis root zone soil moisture (0-100 cm; percentile units)	[-]
PRMC_ANA_PRCNTL	Analysis total profile soil moisture (0-CLSM_DZPR; percentile units)	[-]
TSURF_ANA	Analysis surface temperature	K
TSOIL1_ANA	Analysis first-layer soil temperature	K
SFMC_ANA_MINUS_ENSSTD_PRCNTL	Analysis surface soil moisture minus one ensemble std-dev (percentile units)	[-]
SFMC_ANA_PLUS_ENSSTD_PRCNTL	Analysis surface soil moisture plus one ensemble std-dev (percentile units)	[-]
RZMC_ANA_MINUS_ENSSTD_PRCNTL	Analysis root zone soil moisture minus one ensemble std-dev (percentile units)	[-]
RZMC_ANA_PLUS_ENSSTD_PRCNTL	Analysis root zone soil moisture plus one ensemble std-dev (percentile units)	[-]
PRMC_ANA_MINUS_ENSSTD_PRCNTL	Analysis total profile soil moisture minus one ensemble std-dev (percentile units)	[-]
PRMC_ANA_PLUS_ENSSTD_PRCNTL	Analysis total profile soil moisture plus one ensemble std-dev (percentile units)	[-]
TSURF_ANA_ENSSTD	Ensemble std-dev of TSURF_ANA	K
TSOIL1_ANA_ENSSTD	Ensemble std-dev of TSOIL1_ANA	K

[‡] *Output is* only *stored at times and locations for which input Level 1, 2, or 3 observations are assimilated. If more than one overpass occurs for a given grid cell within the 3 hour window, the latest overpass time prevails.*

[+] *Observed brightness temperatures that originate from 36 km L1C_TB files are posted at 9 km here for convenience (as average over fore and aft brightness temperature if stored separately in L1C_TB product).*

[#] *Model brightness temperatures that correspond to 36 km observations from the L1C_TB product are aggregated from 9 km to 36 km and then posted at 9 km for convenience (Z-scores are computed after aggregation to 36 km).*

[&] *Including flags that originate from Level 1-3 files.*

Table 6b. L4_SM analysis update output. TIME indicates approximate SMAP overpass time and is valid for observed/assimilated (OBS) data and brightness temperature model forecast (FCST) output. For files with 0z time stamp, for example, TIME is between 22:30z and 1:30z. Analysis (ANA) and remaining model forecast (FCST) output is valid at exactly 0z, 3z, 6z, etc. Unless otherwise indicated, model-based outputs are ensemble mean values. Brightness temperatures and related fields (TBHCOMP_* and TBVCOMP_*) are composites of geophysical fields with two different effective resolutions (9 km and 36 km) that are posted on the 9 km EASE grid. For each 9 km grid cell, the flags TBHCOMP_RESFLAG and TBVCOMP_RESFLAG indicate the effective resolution at that location. Z-scores are computed separately for each effective resolution using the appropriate climatological fields (Table 7b).

Data group name	**SMAP Level 4 Surface and Root Zone Soil Moisture land model constants**	
File name	SMAP_L4_SM_modparam_vVVV.h5 (if stored in separate file)	
File format	netcdf4/hdf5	
Characteristics	Constant at native resolution	
Spatial dimensions	Longitude: **3852**, latitude: **1632** (SMAP 9 km EASE grid)	
Variable name	**Description**	**Units**
CLSM_DZSF^{+}	Catchment model: Depth of surface layer	m
CLSM_DZRZ^{+}	Catchment model: Depth of root zone layer	m
CLSM_DZPR	Catchment model: Depth of profile layer ("depth-to-bedrock" in the Catchment model)	m
CLSM_DZTS^{+}	Catchment model: Depth of soil layer associated with TSURF	m
CLSM_DZGT1^{+}	Catchment model: Depth of soil heat diffusion model layer 1	m
CLSM_DZGT2^{+}	Catchment model: Depth of soil heat diffusion model layer 2	m
CLSM_DZGT3^{+}	Catchment model: Depth of soil heat diffusion model layer 3	m
CLSM_DZGT4^{+}	Catchment model: Depth of soil heat diffusion model layer 4	m
CLSM_DZGT5^{+}	Catchment model: Depth of soil heat diffusion model layer 5	m
CLSM_DZGT6^{+}	Catchment model: Depth of soil heat diffusion model layer 6	m
CLSM_POROS	Catchment model: Soil porosity	$m^3 m^{-3}$
CLSM_COND	Catchment model: Soil saturated hydraulic conductivity	$m s^{-1}$
CLSM_PSIS	Catchment model: Matric potential at saturation	m
CLSM_BEE	Catchment model: Clapp-Hornberger parameter	[-]
CLSM_WPWET	Catchment model: Soil wilting point wetness	$m^3 m^{-3}$
CLSM_GNU^{+}	Catchment model: Vertical decay factor for transmissivity	[-]
CLSM_VGWMAX	Catchment model: Maximum amount of water available to vegetation	$kg m^{-2}$
CLSM_VEGCLS	Catchment model: Vegetation class	[-]
CLSM_CDCR1	Catchment model: Catchment deficit at which baseflow ceases	$kg m^{-2}$
CLSM_CDCR2	Catchment model: Maximum water holding capacity of land element	$kg m^{-2}$
CLSM_BF1	Catchment model: Topography-related parameter	$kg^2 m^{-4}$
CLSM_BF2	Catchment model: Topography-related parameter	m
CLSM_BF3	Catchment model: Topography-related parameter	[-]
CLSM_ARS1	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_ARS2	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_ARS3	Catchment model: Topography-related parameter	$kg^{-2} m^4$
CLSM_ARA1	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_ARA2	Catchment model: Topography-related parameter	[-]
CLSM_ARA3	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_ARA4	Catchment model: Topography-related parameter	[-]
CLSM_ARW1	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_ARW2	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_ARW3	Catchment model: Topography-related parameter	$kg^{-2} m^4$
CLSM_ARW4	Catchment model: Topography-related parameter	[-]
CLSM_TSA1	Catchment model: Topography-related parameter	[-]
CLSM_TSA2	Catchment model: Topography-related parameter	[-]
CLSM_TSB1	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_TSB2	Catchment model: Topography-related parameter	$kg^{-1} m^2$
CLSM_ATAU	Catchment model: Topography-related parameter	s
CLSM_BTAU	Catchment model: Topography-related parameter	m

*[**Table 7a**, continues on next page]*

[continued from previous page]

MWRTM_VEGCLS	Microwave RT model: Vegetation class	[-]
MWRTM_SOILCLS	Microwave RT model: Soil class	[-]
MWRTM_POROS	Microwave RT model: Porosity	$m^3 m^{-3}$
MWRTM_SAND	Microwave RT model: Sand fraction	[-]
MWRTM_CLAY	Microwave RT model: Clay fraction	[-]
MWRTM_WANGWT	Microwave RT model: Wang dielectric model transition soil moisture	$m^3 m^{-3}$
MWRTM_WANGWP	Microwave RT model: Wang dielectric mode wilting point soil moisture	$m^3 m^{-3}$
MWRTM_RGHH	Microwave RT model: Roughness parameter	[-]
MWRTM_RGHNRH	Microwave RT model: H-pol. exponent for incidence angle parameterization	[-]
MWRTM_RGHNRV	Microwave RT model: V-pol. exponent for incidence angle parameterization	[-]
MWRTM_RGHPOLMIX	Microwave RT model: Polarization mixing parameter	[-]
MWRTM_OMEGA	Microwave RT model: Single scattering albedo	[-]
MWRTM_BH	Microwave RT model: H-pol. vegetation b parameter	[-]
MWRTM_BV	Microwave RT model: H-pol. vegetation b parameter	[-]
MWRTM_LEWT	Microwave RT model: Parameter to transform LAI into vegetation water content	$kg\ m^{-2}$

[+] *Values are constant in space in the current Catchment model version (may change in future).*

Table 7a. L4_SM time-invariant ancillary input parameters for the Catchment land surface model (CLSM) and the microwave radiative transfer (RT) model (MWRTM). Time-invariant parameters are provided in one file (per L4_SM product version). Soil and vegetation parameters may differ between the Catchment model and the microwave RT model. For example, CLSM_POROS and MWRTM_POROS are not necessarily equal (sections 4.1.2 and 4.1.3).

Data group name	**SMAP Level 4 Surface and Root Zone Soil Moisture climatology**		
File name	SMAP_L4_SM_clim_vVVV.h5 (if stored in separate file)		
File format	netcdf4/hdf5		
Characteristics	Long-term monthly average of instantaneous fields at native resolution		
Spatial dimensions	Longitude: **3852**, latitude: **1632** (SMAP 9 km EASE grid)		
Times	**TBH_*, TBV_*:**	**0z, 3z, 6z, 9z, 12z, 15z, 18z, 21z for each calendar month**	
	SFMC_*, RZMC_*, PRMC_*:	**constant**	
Variable name	**Description**		**Units**
TBH09_L2_SM_AP_MIN	Downscaled (L2_SM_AP) H-pol brightness temperature minimum		K
TBH09_L2_SM_AP_MAX	Downscaled (L2_SM_AP) H-pol brightness temperature maximum		K
TBH09_L2_SM_AP_MEDIAN	Downscaled (L2_SM_AP) H-pol brightness temperature median		K
TBH09_L2_SM_AP_MEAN	Downscaled (L2_SM_AP) H-pol brightness temperature mean		K
TBH09_L2_SM_AP_STD	Downscaled (L2_SM_AP) H-pol brightness temperature std-dev		K
TBH09_MOD_MIN	Catchment model 9 km, 1.41 GHz, H-pol brightness temperature minimum		K
TBH09_MOD_MAX	Catchment model 9 km, 1.41 GHz, H-pol brightness temperature maximum		K
TBH09_MOD_MEDIAN	Catchment model 9 km, 1.41 GHz, H-pol brightness temperature median		K
TBH09_MOD_MEAN	Catchment model 9 km, 1.41 GHz, H-pol brightness temperature mean		K
TBH09_MOD_STD	Catchment model 9 km, 1.41 GHz, H-pol brightness temperature std-dev		K
TBV09_L2_SM_AP_MIN	Downscaled (L2_SM_AP) V-pol brightness temperature minimum		K
TBV09_L2_SM_AP_MAX	Downscaled (L2_SM_AP) V-pol brightness temperature maximum		K
TBV09_L2_SM_AP_MEDIAN	Downscaled (L2_SM_AP) V-pol brightness temperature median		K
TBV09_L2_SM_AP_MEAN	Downscaled (L2_SM_AP) V-pol brightness temperature mean		K
TBV09_L2_SM_AP_STD	Downscaled (L2_SM_AP) V-pol brightness temperature std-dev		K
TBV09_MOD_MIN	Catchment model 9 km, 1.41 GHz, V-pol brightness temperature minimum		K
TBV09_MOD_MAX	Catchment model 9 km, 1.41 GHz, V-pol brightness temperature maximum		K
TBV09_MOD_MEDIAN	Catchment model 9 km, 1.41 GHz, V-pol brightness temperature median		K
TBV09_MOD_MEAN	Catchment model 9 km, 1.41 GHz, V-pol brightness temperature mean		K
TBV09_MOD_STD	Catchment model 9 km, 1.41 GHz, V-pol brightness temperature std-dev		K
TBH36_L1C_TB_MIN^{+}	Observed (L1C_TB) H-pol brightness temperature minimum		K
TBH36_L1C_TB_MAX^{+}	Observed (L1C_TB) H-pol brightness temperature maximum		K
TBH36_L1C_TB_MEDIAN^{+}	Observed (L1C_TB) H-pol brightness temperature median		K
TBH36_L1C_TB_MEAN^{+}	Observed (L1C_TB) H-pol brightness temperature mean		K
TBH36_L1C_TB_STD^{+}	Observed (L1C_TB) H-pol brightness temperature std-dev		K
TBH36_MOD_MIN^{+}	Catchment model 36 km, 1.41 GHz, H-pol brightness temperature min.		K
TBH36_MOD_MAX^{+}	Catchment model 36 km, 1.41 GHz, H-pol brightness temperature max.		K
TBH36_MOD_MEDIAN^{+}	Catchment model 36 km, 1.41 GHz, H-pol brightness temperature median		K
TBH36_MOD_MEAN^{+}	Catchment model 36 km, 1.41 GHz, H-pol brightness temperature mean		K
TBH36_MOD_STD^{+}	Catchment model 36 km, 1.41 GHz, H-pol brightness temperature std-dev		K
TBV36_L1C_TB_MIN^{+}	Observed (L1C_TB) V-pol brightness temperature minimum		K
TBV36_L1C_TB_MAX^{+}	Observed (L1C_TB) V-pol brightness temperature maximum		K
TBV36_L1C_TB_MEDIAN^{+}	Observed (L1C_TB) V-pol brightness temperature median		K
TBV36_L1C_TB_MEAN^{+}	Observed (L1C_TB) V-pol brightness temperature mean		K
TBV36_L1C_TB_STD^{+}	Observed (L1C_TB) V-pol brightness temperature std-dev		K
TBV36_MOD_MIN^{+}	Catchment model 36 km, 1.41 GHz, V-pol brightness temperature min.		K
TBV36_MOD_MAX^{+}	Catchment model 36 km, 1.41 GHz, V-pol brightness temperature max.		K
TBV36_MOD_MEDIAN^{+}	Catchment model 36 km, 1.41 GHz, V-pol brightness temperature median		K
TBV36_MOD_MEAN^{+}	Catchment model 36 km, 1.41 GHz, V-pol brightness temperature mean		K
TBV36_MOD_STD^{+}	Catchment model 36 km, 1.41 GHz, V-pol brightness temperature std-dev		K

*[**Table 7b**, continues on next page]*

[continued from previous page]

SFMC_L2_SM_AP_MIN	Downscaled (L2_SM_AP) surface soil moisture minimum	$m^3 m^{-3}$
SFMC_L2_SM_AP_MAX	Downscaled (L2_SM_AP) surface soil moisture maximum	$m^3 m^{-3}$
SFMC_L2_SM_AP_MEDIAN	Downscaled (L2_SM_AP) surface soil moisture median	$m^3 m^{-3}$
SFMC_L2_SM_AP_MEAN	Downscaled (L2_SM_AP) surface soil moisture mean	$m^3 m^{-3}$
SFMC_L2_SM_AP_STD	Downscaled (L2_SM_AP) surface soil moisture std-dev	$m^3 m^{-3}$
SFMC_L2_SM_AP_CDF_PARAM_0#	SFMC_L2_SM_AP cdf parameter (constant term in 3rd order polynomial)	$m^3 m^{-3}$
SFMC_L2_SM_AP_CDF_PARAM_1#	SFMC_L2_SM_AP cdf parameter (linear term in 3rd order polynomial)	$m^3 m^{-3}$
SFMC_L2_SM_AP_CDF_PARAM_2#	SFMC_L2_SM_AP cdf parameter (quadratic term in 3rd order polynomial)	$m^3 m^{-3}$
SFMC_L2_SM_AP_CDF_PARAM_3#	SFMC_L2_SM_AP cdf parameter (cubic term in 3rd order polynomial)	$m^3 m^{-3}$
SFMC_MOD_MIN	Catchment model surface soil moisture (0-5 cm) minimum	$m^3 m^{-3}$
SFMC_MOD_MAX	Catchment model surface soil moisture (0-5 cm) maximum	$m^3 m^{-3}$
SFMC_MOD_MEDIAN	Catchment model surface soil moisture (0-5 cm) median	$m^3 m^{-3}$
SFMC_MOD_MEAN	Catchment model surface soil moisture (0-5 cm) mean	$m^3 m^{-3}$
SFMC_MOD_STD	Catchment model surface soil moisture (0-5 cm) std-dev	$m^3 m^{-3}$
SFMC_MOD_CDF_PARAM_0#	SFMC_MOD cdf parameter (constant term in 3rd order polynomial)	$m^3 m^{-3}$
SFMC_MOD_CDF_PARAM_1#	SFMC_MOD cdf parameter (linear term in 3rd order polynomial)	$m^3 m^{-3}$
SFMC_MOD_CDF_PARAM_2#	SFMC_MOD cdf parameter (quadratic term in 3rd order polynomial)	$m^3 m^{-3}$
SFMC_MOD_CDF_PARAM_3#	SFMC_MOD cdf parameter (cubic term in 3rd order polynomial)	$m^3 m^{-3}$
RZMC_MOD_MIN	Catchment model root zone soil moisture (0-100 cm) minimum	$m^3 m^{-3}$
RZMC_MOD_MAX	Catchment model root zone soil moisture (0-100 cm) maximum	$m^3 m^{-3}$
RZMC_MOD_MEDIAN	Catchment model root zone soil moisture (0-100 cm) median	$m^3 m^{-3}$
RZMC_MOD_MEAN	Catchment model root zone soil moisture (0-100 cm) mean	$m^3 m^{-3}$
RZMC_MOD_STD	Catchment model root zone soil moisture (0-100 cm) std-dev	$m^3 m^{-3}$
RZMC_MOD_CDF_PARAM_0#	RZMC_MOD cdf parameter (constant term in 3rd order polynomial)	$m^3 m^{-3}$
RZMC_MOD_CDF_PARAM_1#	RZMC_MOD cdf parameter (linear term in 3rd order polynomial)	$m^3 m^{-3}$
RZMC_MOD_CDF_PARAM_2#	RZMC_MOD cdf parameter (quadratic term in 3rd order polynomial)	$m^3 m^{-3}$
RZMC_MOD_CDF_PARAM_3#	RZMC_MOD cdf parameter (cubic term in 3rd order polynomial)	$m^3 m^{-3}$
PRMC_MOD_MIN	Catchment model total profile soil moisture (0-DZPR) minimum	$m^3 m^{-3}$
PRMC_MOD_MAX	Catchment model total profile soil moisture (0-DZPR) maximum	$m^3 m^{-3}$
PRMC_MOD_MEDIAN	Catchment model total profile soil moisture (0-DZPR) median	$m^3 m^{-3}$
PRMC_MOD_MEAN	Catchment model total profile soil moisture (0-DZPR) mean	$m^3 m^{-3}$
PRMC_MOD_STD	Catchment model total profile soil moisture (0-DZPR) std-dev	$m^3 m^{-3}$
PRMC_MOD_CDF_PARAM_0#	PRMC_MOD cdf parameter (constant term in 3rd order polynomial)	$m^3 m^{-3}$
PRMC_MOD_CDF_PARAM_1#	PRMC_MOD cdf parameter (linear term in 3rd order polynomial)	$m^3 m^{-3}$
PRMC_MOD_CDF_PARAM_2#	PRMC_MOD cdf parameter (quadratic term in 3rd order polynomial)	$m^3 m^{-3}$
PRMC_MOD_CDF_PARAM_3#	PRMC_MOD cdf parameter (cubic term in 3rd order polynomial)	$m^3 m^{-3}$

+*36 km brightness temperature climatology is derived at 36 km resolution and posted at 9 km for convenience.*
#*cdf parameters map from percentile to volumetric units.*

Table 7b. L4_SM "climatology" inputs. The climatology information (i) is used for converting brightness temperature into Z-scores (standard normal deviates) prior to data assimilation, (ii) is used for scaling surface soil moisture analysis outputs into the L2_SM_AP climatology, and (iii) enables users to convert L4_SM surface, root zone, and total profile soil moisture from percentile to volumetric units if needed. The climatologies will initially be derived from SMOS brightness temperature observations and from soil moisture retrievals using SMOS observations with SMAP algorithms until a sufficient amount of SMAP data are available to compute SMAP-based climatologies. Brightness temperature statistics are computed using data only at times and locations for which corresponding observations are also available. There is only one "clim" file for each L4_SM product version.

7. References

Andreadis, K. and D. Lettenmaier (2005), Assimilating remotely sensed snow observations into a macroscale hydrology model, *Advances in Water Resources*, *29*, 872–886.

Bartalis Z., W. Wagner, V. Naeimi, S. Hasenauer, K. Scipal, H. Bonekamp, J. Figa, and C. Anderson (2007), Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT), *Geophys. Res. Lett. 34*, L20401, doi:10.1029/2007GL031088.

Bolten, J. D., W. T. Crow, T. J. Jackson, X. Zhan, and C. A. Reynolds (2010), Evaluating the utility of remotely-sensed soil moisture retrievals for operational agricultural drought monitoring, *IEEE Journal of Selected Topics in Earth Observations and Remote Sensing,* 3, 57-66, doi:10.1109/JSTARS.2009.2037163.

Crow, W. T. and R. H. Reichle (2008), Comparison of adaptive filtering techniques for land surface data assimilation, *Water Resources Research, 44*, W08423, doi:10.1029/2008WR006883.

Crow, W. T., and E. Van Loon (2006), Impact of Incorrect Model Error Assumptions on the Sequential Assimilation of Remotely Sensed Surface Soil Moisture, *J. Hydrometeorol.*, 7, 421-432.

Crow, W. T., and E. F. Wood (2003), The assimilation of remotely sensed soil brightness temperature imagery into a land surface model using ensemble Kalman filtering: A case study based on ESTAR measurements during SGP97, *Adv. Water Resour., 26*, 137-149.

Crow, W. T., and X. W. Zhan (2007), Continental-scale evaluation of remotely sensed soil moisture products, *IEEE Trans. Geosci. Remote. Sens., 4*, 451-455, doi:10.1109/LGRS.2007.896533.

Crow, W. T., M. Drusch, and E. F. Wood (2001), An observation system simulation experiment for the impact of land surface heterogeneity on AMSR-E soil moisture retrievals, *IEEE Trans. Geosci. Remote. Sens.*, *39*, 1622–1632.

Crow, W. T., T. Chen, D. Entekhabi, P. Houser, A. Hsu, T. Jackson, E. Njoku, P. O'Neill, J. Shi, and X. Zhan (2005a), An observing system simulation experiment for Hydros radiometer-only soil moisture products, *IEEE Trans. Geosci. Remote Sens.*, 43, 1289– 1303.

Crow, W. T., R. D. Koster, R. H. Reichle, and H. O. Sharif (2005b), Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products, *Geophys. Res. Lett.*, *32* (24), Art. No. L24405, doi:10.1029/2005GL024889.

Crow, W. T., D. G. Miralles and M. H. Cosh (2010), A quasi-global evaluation system for satellite-based surface soil moisture retrievals, *IEEE Transactions on Geoscience and Remote Sensing, 48*, 2516-2527, 10.1109/IGARSS.2008.4779051.

De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, V. R. N. Pauwels, and N. E. C. Verhoest (2007), Correcting for Forecast Bias in Soil Moisture Assimilation with the Ensemble Kalman Filter, *Water Resources Research, 43*, W09410, doi:10.1029/2006WR005449.

De Lannoy, G. J. M., R. H. Reichle, P. R. Houser, K. R. Arsenault, N. E. C. Verhoest, and V. R. N. Pauwels (2010), Satellite-Scale Snow Water Equivalent Assimilation into a High-Resolution Land Surface Model, *Journal of Hydrometeorology, 11*, 352-369, doi:10.1175/2009JHM1192.1.

De Lannoy, G. J. M., R. H. Reichle, K. R. Arsenault, P. R. Houser, S. V. Kumar, N. E. C. Verhoest, and V. R. N. Pauwels (2011), Multi-Scale Assimilation of AMSR-E Snow Water Equivalent and MODIS Snow Cover Fraction in Northern Colorado, *Water Resources Research*, *48*, W01522, doi:10.1029/2011WR010588..

De Lannoy, G. J. M., R. H. Reichle, and V. R. N. Pauwels (2012), Global Calibration of the GEOS-5 L-band Microwave Radiative Transfer Model over Land Using SMOS Observations, submitted.

De Rosnay, P., M. Drusch, A. Boone, G. Balsamo, B. Decharme, P. Harris, Y. Kerr, T. Pellarin, J. Polcher, and J.-P. Wigneron (2009), AMMA Land Surface Model Intercomparison Experiment coupled to the Community Microwave Emission Model: ALMIP-MEM. *J. Geophys. Res., 114*, D05108, doi:10.1029/2008JD010724.

Dee, D. P. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation, *Mon. Wea. Rev.*, *123*, 1128-1145.

Dee, D. P. (2005), Bias and data assimilation, *Quart. J. R. Meteo. Soc., 131* (613), 3323-3343 Part C, doi:10.1256/qj.05.137.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Quart. J. R. Meteo. Soc., 131* (613), 3385-3396 Part C, doi:10.1256/qj.05.108.

Dorigo, W.A., P. Van Oevelen, W. Wagner, M. Drusch, S. Mecklenburg, A. Robock, and T. Jackson (2011), A New International Network for in Situ Soil Moisture Data, *EOS, Transactions*, *American Geophysical Union*, 92, 141–142.

Draper, C. S., R. H. Reichle, G. J. M. De Lannoy, and Q. Liu (2012), Assimilation of passive and active microwave soil moisture retrievals, *Geophysical Research Letters, 39*, L04401, doi:10.1029/2011GL050655.

Drusch, M. (2007), Initializing numerical weather prediction models with satellite derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set, *J. Geophys. Res.*, *112*, D03102, doi:10.1029/2006JD007478.

Drusch, M., E. F. Wood, and H. Gao (2005), Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture, *Geophys. Res. Let.*, *32*, L15403, doi:10.1029/2005GL023623.

Drusch, M., T. Holmes, P. de Rosnay, and G. Balsamo (2009), Comparing ERA-40 based L-band brightness temperatures with Skylab observations: a calibration/validation study using the Community Microwave Emission Model, *J. Hydrometeorol.*, *10*, 213–226, doi:10.1175/2008JHM964.1.

Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model, 2: Parameter estimation and model demonstration, *J. Geophys. Res. 105*(20), 24823-24838.

Dunne, S., and D. Entekhabi (2006), Land surface state and flux estimation using the ensemble Kalman smoother during the Southern Great Plains 1997 field experiment, *Water Resour. Res.*, *42*, W01407, doi:10.1029/2005WR004334.

Durand, M., and S. Margulis (2008), Effects of uncertainty magnitude and accuracy on assimilation of multi-scale measurements for snowpack characterization, *J. Geophys. Res.*, *113*, D02105, doi:10.1029/2007JD008662.

Entekhabi, D., H. Nakamura, and E. G. Njoku (1994), Solving the inverse problems for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely-sensed observations, , *IEEE Trans. Geosci. Remote Sens., 32*, 438-448.

Entekhabi, D., E. Njoku, P. Houser, M. Spencer, T. Doiron, J. Smith, R. Girard, S. Belair, W. Crow, T. Jackson, Y. Kerr, J. Kimball, R. Koster, K. McDonald, P. O'Neill, T. Pultz, S. Running, J. C. Shi, E. Wood, and J. van Zyl (2004), The Hydrosphere State (HYDROS) mission concept: An Earth system pathfinder for global mapping of soil moisture and land freeze/thaw, *IEEE Trans. on Geosci. and Remote Sensing*, *42*, 2184-2195.

Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow (2010a), Performance Metrics for Soil Moisture Retrievals and Application Requirements, *Journal of Hydrometeorology, 11*, 832-840, doi:10.1175/2010JHM1223.1.

Entekhabi, D., and Coauthors (2010b), The Soil Moisture Active and Passive (SMAP) Mission, *Proceedings of the IEEE*, *98*, 704-716, doi:10.1109/JPROC.2010.2043918.

Evensen, G. (2003), The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics, 53*, 343-367, doi:10.1007/s10236-003-0036-9.

Gao, H., E. F. Wood, T. J. Jackson, M. Drusch, and R. Bindlish (2006), Using TRMM/TMI to Retrieve Surface Soil Moisture over the Southern United States from 1998 to 2002, *J. Hydrometeorol.*, *7*, 23-38.

Hoeben, R., and P. A. Troch (2000), Assimilation of active microwave observation data for soil moisture profile estimation, *Water Resources Research, 36,* 2805-2819.

Holmes, T. R. J., T. J. Jackson, R. H. Reichle, and J. Basara (2011), An Assessment of Surface Soil Temperature Products from Numerical Weather Prediction Models Using Ground-based Measurements, *Water Resources Research*, *48*, W02531, doi:10.1029/2011WR010538.

International GEWEX Project Office (2002), *The Second Global Soil Wetness Project science and implementation plan,* IGPO Publication Series 37, 69 pp.

Jackson, T., and T. Schmugge (1991), Vegetation effects on the microwave emission of soils, *Remote Sens. Environ.*, *36*, 203– 212.

Jackson, T. J., M. Cosh, R. Bindlish, P. Starks, D. Bosch, M. Seyfried, D. Goodrich, S. Moran, and J. Du (2010), Validation of Advanced Microwave Scanning Radiometer Soil Moisture Products, *IEEE Trans. Geosci. Remote Sens., 48*, 4256-4272, doi: 10.1109/TGRS.2010.2051035.

Jones, A. S., T. Vukicevic, and T. H. Vonder Haar (2004), A microwave satellite observational operator for variational data assimilation of soil moisture, *Journal of Hydrometeorology, 5*, 213-229.

Keppenne, C. L. (2000), Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter, *Mon. Wea. Rev., 128*, 1971–1981.

Kerr, Y. H., P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger (2001), Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission, *IEEE Trans. Geosci. Remote Sens. 39*(8), 1729-1735.

Kerr, Y., and Coauthors (2010), The SMOS mission: New tool for monitoring key elements of the global water cycle, *Proceedings of the IEEE, 98*, 666-687.

Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar (2000a), A catchment-based approach to modeling land surface processes in a general circulation model, 1: Model structure, *J. Geophys. Res. 105*(20), 24809-24822.

Kumar, S. V., R. H. Reichle, R. D. Koster, W. T. Crow, and C. D. Peters-Lidard (2009), Role of subsurface physics in the assimilation of surface soil moisture observations, *Journal of Hydrometeorology*, *10*, 1534-1547, doi:10.1175/2009JHM1134.1.

Kumar, S. V., C. Peters-Lidard, Y. Tian, R. H. Reichle, C. Alonge, J. Geiger, J. Eylander, and P. Houser (2008b), An integrated hydrologic modeling and data assimilation framework enabled by the Land Information System (LIS), *IEEE Computer, 41*, 52-59, doi:10.1109/MC.2008.511.

Kumar, S. V., R. H. Reichle, C. D. Peters-Lidard, R. D. Koster, X. Zhan, W. T. Crow, J. B. Eylander, and P. R. Houser (2008a), A Land Surface Data Assimilation Framework using the Land Information System: Description and Applications, *Advances in Water Resources, 31*, 1419-1432, doi:10.1016/j.advwatres.2008.01.013.

Kumar, S. V., R. H. Reichle, K. W. Harrison, C. D. Peters-Lidard, S. Yatheendradas, and J. A. Santanello (2011), A comparison of methods for a priori bias correction in soil moisture data assimilation, *Water Resources Research*, *48*, W03515, doi:10.1029/2010WR010261.

Lagerloef, G. S. E., F. R. Colomb, D. M. Le Vine, F. J. Wentz, S. H. Yueh, C. S. Ruf, J. Lilly, J. Gunn, Y. Chao, A. de Charon, G. Feldman, and C. T. Swift (2008), The Aquarius/SAC-D mission, *Oceanography, 21*, 68-81.

Larson K. M., E. E. Small, E. Gutmann, A. Bilich, P. Axelrad, and J. Braun (2008), Using GPS multipath to measure soil moisture fluctuations: initial results, *GPS Solutions, 12,* 173-177, doi:10.1007/s10291-007-0076-6.

Li, L., P. Gaiser, T. Jackson, R. Bindlish, and J. Du (2007), WindSat soil moisture algorithm and validation, *IEEE International Geoscience and Remote Sensing Symposium, 23-28 July 2007*, 1188-1191, doi:10.1109/IGARSS.2007.4423017.

Liu, Q., R. H. Reichle, R. Bindlish, M. H. Cosh, W. T. Crow, R. de Jeu, G. J. M. De Lannoy, G. J. Huffman, and T. J. Jackson (2011), The contributions of precipitation and soil moisture observations to the skill of soil moisture estimates in a land data assimilation system, *Journal of Hydrometeorology*, *12*, 750-765, doi:10.1175/JHM-D-10-05000.

Maggioni, V., R. H. Reichle, and E. N. Anagnostou (2011), The effect of satellite-rainfall error modeling on soil moisture prediction uncertainty, *Journal of Hydrometeorology*, *12*, 413-428, doi:10.1175/2011JHM1355.1.

Maggioni, V., R. H. Reichle, and E. N. Anagnostou (2012a), The impact of rainfall error characterization on the estimation of soil moisture fields in a land data assimilation system, *Journal of Hydrometeorology*, *13*, 1107-1118, doi:10.1175/JHM-D-11-0115.1.

Maggioni, V., R. H. Reichle, and E. N. Anagnostou (2012b), The efficiency of assimilating satellite soil moisture retrievals in a land data assimilation system using different rainfall error models, *Journal of Hydrometeorology*, submitted.

Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne (2002), Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment, *Water Resour. Res.*, *38*, 1299, doi:10.1029/2001WR001114.

Margulis, S.A., and D. Entekhabi (2003), Variational Assimilation of Radiometric Surface Temperature and Reference-Level Micrometeorology into a Model of the Atmospheric Boundary Layer and Land Surface, *Mon. Wea. Rev.*, *131*, 1272–1288.

McLaughlin, D., Y. H. Zhou, D. Entekhabi, and V. Chatdarong (2006), Computational issues for large-scale land surface data assimilation problems, *Journal of Hydrometeorology, 7*, 494-510.

Mitchell K. E., et al. (2004), The multi-institution North American Land Data Assimilation System (NLDAS): Utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system, *J. Geophys. Res., 109,* D07S90, doi: 10.1029/2003JD003823.

Njoku, E. G. (2011), Updated daily, AMSR-E/Aqua L2B surface soil moisture, ancillary parms, & QC EASE-Grids, June 2002 to June 2011, Boulder, CO, USA: National Snow and Ice Data Center (http://nsidc.org/data/amsre/), Digital media.

Owe, M., R. de Jeu, and T. Holmes (2008), Multisensor historical climatology of satellite-derived global land surface moisture, *J. Geophys. Res., 113*, F01002, doi:10.1029/2007JF000769.

Pan, M., and E. F. Wood (2006), Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter, *J. Hydrometeorol., 7*, 534-547.

Pan, M., E. F. Wood, R. Wojcik, and M. F. McCabe (2008), Estimation of regional terrestrial water cycle using multi-sensor remote sensing observations and data assimilation, *Remote Sensing of Environment, 112,* 1282-1294, doi:10.1016/j.rse.2007.02.039.

Pellarin, T., et al. (2003), Two-year global simulation of L-band brightness temperature over land, *IEEE Trans. Geosci. Remote Sens., 41*, 2135–2139.

Reichle, R. H. (2008), Data Assimilation Methods in the Earth Sciences, *Advances in Water Resources, 31*, 1411-1418, doi:10.1016/j.advwatres.2008.01.001.

Reichle, R. H., and R. D. Koster (2003), Assessing the impact of horizontal error correlations in background fields on soil moisture estimation, *J. Hydrometeorol.*, *4* (6), 1229-1242.

Reichle, R. H., and R. D. Koster (2004), Bias reduction in short records of satellite soil moisture, *Geophys. Res. Let.*, *31*, doi:10.1029/2004GL020938.

Reichle, R. H., and R. D. Koster (2005), Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model, *Geophys. Res. Let.*, *32*, doi:10.1029/2004GL021700.

Reichle, R. H., D. Entekhabi, and D. B. McLaughlin (2001), Downscaling of radiobrightness measurements for soil moisture estimation: A four-dimensional variational data assimilation approach, *Water Resources Research, 37*, 2353-2364, doi:10.1029/2001WR000475.

Reichle, R. H., D. McLaughlin, and D. Entekhabi (2002a), Hydrologic data assimilation with the Ensemble Kalman filter, *Mon. Weather Rev. 130*(1), 103-114.

Reichle, R. H., J. P. Walker, R. D. Koster, and P. R. Houser (2002b), Extended versus Ensemble Kalman filtering for land data assimilation, *J. Hydrometeorol.* 3(6), 728-740.

Reichle, R. H., R. D. Koster, J. Dong, and A. A. Berg (2004), Global soil moisture from satellite observations, land surface models, and ground data: Implications for data assimilation, *J. Hydrometeorol.*, *5*, 430-442.

Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe (2007), Comparison and assimilation of global soil moisture retrievals from AMSR-E and SMMR, *J. Geophys. Res*., 112, D09108, doi:10.1029/2006JD008033.

Reichle, R. H., W. T. Crow, R. D. Koster, H. Sharif, and S. P. P. Mahanama (2008a), Contribution of soil moisture retrievals to land data assimilation products, *Geophysical Research Letters, 35*, L01404, doi:10.1029/2007GL031986.

Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008b), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resources Research, 44*, W03423, doi:10.1029/2007WR006357.

Reichle, R. H., M. G. Bosilovich, W. T. Crow, R. D. Koster, S. V. Kumar, S. P. P. Mahanama, and B. F. Zaitchik (2009a), Recent Advances in Land Data Assimilation at the NASA Global Modeling and Assimilation Office, *In: Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, edited by Seon K. Park and Liang Xu, 407-428, Springer Verlag, New York, doi:10.1007/978-3-540-71056-1.

Reichle, R. H., W. Crow, R. D. Koster, and J. Kimball (2009b), External Review of the SMAP L4_SM Algorithm and Product, *SMAP Science Document no: 031, version 1.0*, 50 pages.

Reichle, R. H., S. V. Kumar, S. P. P. Mahanama, R. D. Koster, and Q. Liu (2010), Assimilation of satellite-derived skin temperature observations into land surface models, Journal of Hydrometeorology, *Journal of Hydrometeorology, 11*, 1103-1122, doi:10.1175/2010JHM1262.1.

Reichle, R. H., R. D. Koster, G. J. M. De Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Toure (2011), Assessment and enhancement of MERRA land surface hydrology estimates, *Journal of Climate*, *24*, 6322-6338, doi:10.1175/JCLI-D-10-05033.1.

Rienecker, M. M., M. J. Suarez, R. Todling, J. Bacmeister, L. Takacs, H.-C. Liu, W. Gu, M. Sienkiewicz, R. D. Koster, R. Gelaro, I. Stajner, and J. E. Nielsen (2008), The GEOS-5 Data Assimilation System - Documentation of Versions 5.0.1, 5.1.0, and 5.2.0. *Technical Report Series on Global Modeling and Data Assimilation*, *27,* http://gmao.gsfc.nasa.gov/pubs/docs/Rienecker369.pdf.

Rodell, M., and P. R. Houser (2004), Updating a land surface model with MODIS-derived snow cover, *Journal of Hydrometeorology, 5,* 1064-1075.

Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, and D. L. Toll (2003), The Global Land Data Assimilation System, *Bull. Amer. Meteorol. Soc.*, *85*, 381-394, doi:10.1175/BAMS-85-3-381.

Rodell, M., P.R. Houser, A.A. Berg, and J.S. Famiglietti (2005), Evaluation of 10 Methods for Initializing a Land Surface Model, *J. Hydrometeorology, 6,* 146–155.

Seuffert, G., H. Wilker, P. Viterbo, J.-F. Mahfouf, M. Drusch, and J.-C. Calvet (2003), Soil moisture analysis combining screen-level parameters and microwave brightness temperature: A test with field data, *Geophys. Res. Lett.*, 30, 1498, doi:10.1029/2063GL017128.

Sabater, J. M., P. De Rosnay, and G. Balsamo (2011), Sensitivity of L-band NWP forward modelling to soil roughness, International Journal of Remote Sensing, doi:10.1080/01431161.2010.507260.

Scipal, K., M. Drusch, and W. Wagner (2008), Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weal-her prediction system, *Advances in Water Resources, 31,* 1101-1112, doi:10.1016/j.advwatres.2008.04.013.

Space Studies Board (2007), *Earth Science and Applications From Space: National Imperatives for the Next Decade and Beyond*, 400 pp., Natl. Acad. of Sci., Washington, DC.

Wagner, W., V. Naeimi, K. Scipal, R. de Jeu, and J. Martínez-Fernández (2007), Soil moisture from operational meteorological satellites, *Hydrogeology Journal, 15,* 121-131.

Walker, J. P., and P. R. Houser (2001), A methodology for intializing soil moisture in a global climate model: Assimilation of near-surface soil moisture observations, *J. Geophys. Res., 106*, 11,761-11,774.

Wang, J. R., and B. J. Choudhury (1981), Remote sensing of soil moisture content over bare field at 1.4 GHz frequency, *J. Geophys. Res., 86*, 5277-5282.

Wang, J. R., and T. Schmugge (1980), An empirical model for the complex dielectric permittivity of soils as a function of water content, *IEEE Trans. Geosci. Rem. Sens.*, *18*, 288-295.

Yi, Y., J. Kimball, L. Jones, R. H. Reichle, and K. McDonald (2011), Evaluation of MERRA land surface estimates in preparation for the Soil Moisture Active Passive mission, *Journal of Climate*, *24*, 3797–3816, doi:10.1175/2011JCLI4034.1.

Zaitchik, B.F., and M. Rodell (2009), Forward-Looking Assimilation of MODIS-Derived Snow-Covered Area into a Land Surface Model, *J. Hydrometeor.*, *10*, 130–148, doi:10.1175/2008JHM1042.1.

Zaitchik, B. F., M. Rodell, and R. H. Reichle (2008), Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin, *Journal of Hydrometeorology, 9*, 535-548, doi:10.1175/2007JHM951.1.

Zhan, X. W., P. R. Houser, J. P. Walker, and W. T. Crow (2006), A method for retrieving high-resolution surface soil moisture from hydros L-band radiometer and radar observations, *IEEE Trans. Geosci.* Remote Sens., *44*, 1534-1544.

Zhou, Y., D. McLaughlin, and D. Entekhabi (2006), Assessing the Performance of the Ensemble Kalman Filter for Land Surface Data Assimilation, *Mon. Weather Rev. 134*, 2128-2142.

Zweck, C., M. Zreda, J. Shuttleworth, and X. Zeng (2008), COSMOS: COsmic-ray Soil Moisture Observing System planned for the United States, American Geophysical Union, Fall Meeting 2008, abstract #H41D-0909.

8. SMAP Reference Documents

Requirements:

- SMAP Level 1 Mission Requirements and Success Criteria. (Appendix O to the Earth Systematic Missions Program Plan: Program-Level Requirements on the Soil Moisture Active Passive Project.). NASA Headquarters/Earth Science Division, Washington, DC.
- SMAP Level 2 Science Requirements. SMAP Project, JPL D-45955, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Level 3 Science Algorithms and Validation Requirements. SMAP Project, JPL D-45993, Jet Propulsion Laboratory, Pasadena, CA.

Plans:

- SMAP Science Data Management and Archive Plan. SMAP Project, JPL D-45973, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Science Data Calibration and Validation Plan. SMAP Project, JPL D-52544, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Applications Plan. SMAP Project, JPL D-53082, Jet Propulsion Laboratory, Pasadena, CA.

ATBDs:

- SMAP Algorithm Theoretical Basis Document: L1B and L1C Radar Products. SMAP Project, JPL D-53052, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L1B Radiometer Product. SMAP Project, GSFC-SMAP-006, NASA Goddard Space Flight Center, Greenbelt, MD.
- SMAP Algorithm Theoretical Basis Document: L1C Radiometer Product. SMAP Project, JPL D-53053, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L2 & L3 Radar Soil Moisture (Active) Products. SMAP Project, JPL D-66479, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L2 & L3 Radiometer Soil Moisture (Passive) Products. SMAP Project, JPL D-66480, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L2 & L3 Radar/Radiometer Soil Moisture (Active/Passive) Products. SMAP Project, JPL D-66481, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L3 Radar Freeze/Thaw (Active) Product. SMAP Project, JPL D-66482, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L4 Surface and Root Zone Soil Moisture Product. SMAP Project, JPL D-66483, Jet Propulsion Laboratory, Pasadena, CA.
- SMAP Algorithm Theoretical Basis Document: L4 Carbon Product. SMAP Project, JPL D-66484, Jet Propulsion Laboratory, Pasadena, CA.

Ancillary Data Reports:

- Ancillary Data Report: Crop Type. SMAP Project, JPL D-53054, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Digital Elevation Model. SMAP Project, JPL D-53056, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Land Cover Classification. SMAP Project, JPL D-53057, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Soil Attributes. SMAP Project, JPL D-53058, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Static Water Fraction. SMAP Project, JPL D-53059, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Urban Area. SMAP Project, JPL D-53060, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Vegetation Water Content. SMAP Project, JPL D-53061, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Permanent Ice. SMAP Project, JPL D-53062, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Precipitation. SMAP Project, JPL D-53063, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Snow. SMAP Project, GSFC-SMAP-007, NASA Goddard Space Flight Center, Greenbelt, MD.
- Ancillary Data Report: Surface Temperature. SMAP Project, JPL D-53064, Jet Propulsion Laboratory, Pasadena, CA.
- Ancillary Data Report: Vegetation and Roughness Parameters. SMAP Project, JPL D-53065, Jet Propulsion Laboratory, Pasadena, CA.

Other Documents:

- Reichle, R. H., W. Crow, R. D. Koster, and J. Kimball (2009b), External Review of the SMAP L4_SM Algorithm and Product, *SMAP Science Document no: 031, version 1.0*, 50 pages.

9. List of Acronyms and Abbreviations

AMSR-E	Advanced Microwave Scanning Radiometer for EOS
ARS	Agricultural Research Service
ASCAT	Advanced Scatterometer
ATBD	Algorithm Theoretical Basis Document
Cal/Val	Calibration and validation
cdf	Cumulative distribution function
CEOP	Coordinated Energy and Water Cycle Observations Project
CLSM	Catchment Land Surface Model
CMAP	CPC Merged Analysis of Precipitation
CPC	Climate Prediction Center
CPU	Central Processing Unit
DAAC	Distributed Active Archive Center
EASE	Equal-Area Scalable Earth (grid)
ECMWF	European Centre for Medium-Range Weather Forecasting
EnKF	Ensemble Kalman filter
EOS	Earth Observing System
ERS	European Remote Sensing (satellite)
GEOS-5	Goddard Earth Observing System Model, Version 5
GEWEX	Global Energy and Water Cycle Experiment
GLDAS	Global Land Data Assimilation System
GMAO	Global Modeling and Assimilation Office
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Measurement (mission)
GSFC	Goddard Space Flight Center
HQ	Headquarters
IOC	In-orbit checkout (period)
ISCCP	International Satellite Cloud Climatology Project
L1C_TB	SMAP Level 1 Radiomter Brightness Temperature (Half-Orbit) (product/algorithm)
L2_SM_AP	SMAP Level 2 Radar and Radiometer Soil Moisture (product/algorithm)
L2_SM_P	SMAP Level 2 Radiometer Soil Moisture (product/algorithm)
L3_FT_A	SMAP Level 3 Radar Freeze-Thaw (product/algorithm)
L4_C	SMAP Level 4 Carbon Net Ecosystem Exchange (product/algorithm)
L4_SM	SMAP Level 4 Surface and Root Zone Soil Moisture (product/algorithm)
MERRA	Modern-Era Retrospective analysis for Research and Applications
MWRTM	Microwave Radiative Transfer Model
NASA	National Aeronautics and Space Administration
NCCS	NASA Center for Climate Simulation
NESDIS	National Environmental Satellite, Data, and Information Service
NOAA	National Oceanic and Atmospheric Administration

NRC	National Research Council
OSSE	Observing System Simulation Experiment
QC	Quality control
RFI	Radio-frequency interference
RMSE	Root-Mean-Square Error
SCAN	Soil Climate Analysis Network
SMAP	Soil Moisture Active Passive (mission)
SMMR	Scanning Multichannel Microwave Radiometer
SMOS	Soil Moisture Ocean Salinity (mission)
SWE	Snow water equivalent
TMI	TRMM Microwave Imager
TRMM	Tropical Rainfall Measurement Mission
USDA	United States Department of Agriculture
USGS	United States Geological Survey

10. Glossary

Adapted from: Earth Observing System Data and Information System (EOSDIS) Glossary
http://www-v0ims.gsfc.nasa.gov/v0ims/DOCUMENTATION/GLOS-ACR/glossary.of.terms.html.

ALGORITHM. (1) Software delivered by a science investigator to be used as the primary tool in the generation of science products. The term includes executable code, source code, job control scripts, as well as documentation. (2) A prescription for the calculation of a quantity; used to derive geophysical properties from observations and to facilitate calculation of state variables in models.

ANCILLARY DATA. Data other than instrument data required to perform an instrument's data processing. They include orbit data, attitude data, time information, spacecraft engineering data, calibration data, data quality information, data from other instruments (spaceborne, airborne, ground-based) and models.

BROWSE. A representation of a data set or data granule used to pre-screen data as an aid to selection prior to ordering. A data set, typically of limited size and resolution, created to rapidly provide an understanding of the type and quality of available full resolution data sets. It may also enable the selection of intervals for further processing or analysis of physical events. For example, a browse image might be a reduced resolution version of a single channel from a multi-channel instrument. Note: Full resolution data sets may be browsed.

BROWSE DATA PRODUCT. Subsets of a larger data set, generated for the purpose of allowing rapid interrogation (i.e., browse) of the larger data set by a potential user. For example, the browse product for an image data set with multiple spectral bands and moderate spatial resolution might be an image in two spectral channels, at a degraded spatial resolution. The form of browse data is generally unique for each type of data set and depends on the nature of the data and the criteria used for data selection within the relevant scientific disciplines.

Dynamic Browse. Refers to the generation of a browse product, including subsetting and/or resampling of data, by command of the user engaged in the browse activity. The browse data set is built in real-time, or near-real-time, as part of the browse activity.

Static Browse. Refers to interrogation of browse products which have been generated (through subsetting and/or resampling) before any user browses that particular data set.

CALIBRATION. (1) The activities involved in adjusting an instrument to be intrinsically accurate, either before or after launch (i.e., "instrument calibration"). (2) The process of collecting instrument characterization information (scale, offset, nonlinearity, operational, and environmental effects), using either laboratory standards, field standards, or modeling, which is used to interpret instrument measurements (i.e., "data calibration").

CALIBRATION DATA. The collection of data required to perform calibration of the instrument science and engineering data, and the spacecraft or platform engineering data. It includes pre-flight calibrator measurements, calibration equation coefficients derived from calibration software routines, and ground truth data that are to be used in the data calibration processing routine.

CORRELATIVE DATA. Scientific data from other sources used in the interpretation or validation of instrument data products, e.g. ground truth data and/or data products of other instruments. These data are not utilized for processing instrument data.

DATA PRODUCT. A collection (1 or more) of parameters packaged with associated ancillary and labeling data. Uniformly processed and formatted. Typically uniform temporal and spatial resolution. (Often the collection of data distributed by a data center or subsetted by a data center for distribution.) There are two types of data products:

Standard - A data product produced by a community concensus algorithm. Typically produced for a wide community. May be produced routinely or on-demand. If produced routinely, typically produced over most or all of the available independent variable space. If produced on-demand, produced only on request from users for particular research needs typically over a limited range of independent variable space.

Special - A data product produced by a research status algorithm. May migrate to a community consesus algorithm at a later time. If adequate community interest, may be archived and distibuted by a DAAC.

DATA PRODUCT LEVEL. Data levels 1 through 4 as designated in the EOSDIS Product Type and Processing Level Definitions document.

Raw Data - Data in their original packets, as received from the observer, unprocessed.

Level 0 - Raw instrument data at original resolution, time ordered, with duplicate packets removed.

Level 1A - Reconstructed unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (i.e., platform ephemeris) computed and appended, but not applied to Level 0 data.

Level 1B - Radiometrically corrected and geolocated Level 1A data that have been processed to sensor units.

Level 1C - Level 1B data that have been spatially resampled.

Level 2 - Derived geophysical parameters at the same resolution and location as the Level 1 (1B or 1C) data.

Level 3 - Geophysical or sensor parameters that have been spatially and/or temporally re-sampled (i.e., derived from Level 2 or Level 1 data).

Level 4 - Model output and/or results of lower level data that are not directly derived by the instruments.

DISTRIBUTED ACTIVE ARCHIVE CENTER (DAAC). An EOSDIS facility that archives, and distributes data products, and related information. An EOSDIS DAAC is managed by an institution such as a NASA field center or a university, under terms of an agreement with NASA. Each DAAC contains functional elements for archiving and disseminating data, and for user services and information management. Other (non-NASA) agencies may share management and funding responsibilities for the active archives under terms of agreements negotiated with NASA.

GRANULE. The smallest aggregation of data which is independently managed (i,e., described, inventoried, retrievable). Granules may be managed as logical granules and/or physical granules.

GUIDE. A detailed description of a number of data sets and related entities, containing information suitable for making a determination of the nature of each data set and its potential usefulness for a specific application.

INSTRUMENT DATA. Data specifically associated with the instrument, either because they were generated by the instrument or included in data packets identifed with that instrument. These data consist of instrument science and engineering data, and possible ancillary data.

Instrument Engineering Data. Data produced by the engineering sensor(s) of an instrument that is used to determine the physical state of an instrument in order to operate it, monitor its health, or aid in processing its science data.

Instrument Science Data. Data produced by the science sensor(s) containing the primary observables of an instrument, usually constituting the mission of that instrument.

METADATA. (1) Information about a data set which is provided by the data supplier or the generating algorithm and which provides a description of the content, format, and utility of the data set. Metadata provide criteria which may be used to select data for a particular scientific investigation. (2) Information describing a data set, including data user guide, descriptions of the data set in directories, and inventories, and any additional information required to define the relationships among these.

NEAR REAL-TIME DATA. Data from the source that are available for use within a time that is short in comparison to important time scales in the phenomena being studied.

ORBIT DATA. Data that represent spacecraft locations. Orbit (or ephemeris) data include: Geodetic latitude, longitude and height above an adopted reference ellipsoid (or distance from the center of mass of the Earth); a corresponding statement about the accuracy of the position and the corresponding time of the position (including the time system); some accuracy requirements may be hundreds of meters while other may be a few centimeters.

PARAMETER. A measurable or derived variable represented by the data (e.g. air temperature, snow depth, relative humidity.

QUICK-LOOK DATA. Data available for examination within a short time of receipt, where completeness of processing is sacrificed to achieve rapid availability.

RAW DATA. Numerical values representing the direct observations output by a measuring instrument transmitted as a bit stream in the order they were obtained. (Also see DATA PRODUCT LEVEL.)

REAL-TIME DATA. Data that are acquired and transmitted immediately to the ground (as opposed to playback data). Delay is limited to the actual time (propagation delays) required to transmit the data.

SPACECRAFT ENGINEERING DATA. Data produced by the engineering sensor(s) of a spacecraft that are used to determine the physical state of the spacecraft, in order to operate it or monitor its health.